

On Probabilistic Parametric Inference

Tomaž Podobnik^{1,2} and Tomi Živko²

¹Faculty of Mathematics and Physics, University of Ljubljana and ²Jožef Stefan Institute, Ljubljana, Slovenia

Abstract. An objective operational theory of probabilistic parametric inference is formulated without invoking the so-called non-informative prior probability distributions.

AMS 2000 subject classifications: Primary 62F15; secondary 60B15, 62F25.

Key words and phrases: Inverse probability distributions, non-informative priors, consistency factors, invariant parametric families, interpretations of probability distributions.

1. INTRODUCTION

We make a probabilistic inference about a parameter of a family of the so-called direct probability distributions by specifying a probability distribution that corresponds to the distribution of our belief in different values of the parameter (Jeffreys (1957), § 2.0, p. 22). The probabilistic parametric inference is characteristic of *Bayesian schools* of statistical inference (as opposed to *frequentist schools*), where the name Bayesian is due to the central role of Bayes' Theorem in the process of inference. In the Bayesian paradigms, it is also possible to make statements concerning the values of the inferred parameters in the absence of data, and these statements can be summarized in the so-called (*non-informative*) *prior probability distributions*, (Villegas (1981); see also, for example, Jeffreys (1961), § 1.4, p. 33 and § 3.1, pp. 117-118; Ferguson (1967), § 1.6, pp. 30-31; Berger (1980), § 1.2, pp. 4-5; Rao (1993), § 3.5, p. 86; O'Hagan (1994), § 1.21, p. 23; Kass and Wasserman (1996); Lad (1996), § 3.4, p. 150; Shao (1999), § 4.1.1, p. 193; Robert (2001), § 3.5, pp. 127-140; Casella and Berger (2002), § 7.2.3, p. 324; Jaynes (2003), § 4.1, pp. 87-88; Harney (2003), § 2.1, p. 9; Hogg et al. (2005), § 11.2.1, pp. 583-584). The non-informative prior distributions provide a formal way of expressing ignorance about the inferred parameter (Jeffreys (1961), § 3.1, pp. 117-118; Kass and Wasserman (1996), § 4.1, p. 1355). It has

been asserted (Jeffreys (1957), § 2.3, p. 31; Jeffreys (1961), § 1.5, pp. 36-37; Bernardo (1979), § 5.1, p. 123; Kass and Wasserman (1996), § 4.1, pp. 1355-1356; Robert (2001), § 3.5, p. 127) that there is no objective, unique non-informative prior distribution that represents ignorance. Instead, the priors should be chosen by public agreement, much like units of length and weight, upon which everyone could fall back when the prior information about the inferred parameter is missing.

In the present article, a theory of probabilistic parametric inference is developed without invoking the non-informative prior probability distributions. Moreover, it is demonstrated that the non-informative prior probability distributions necessarily lead to inconsistencies. Sections 2–4 are devoted to formulation of a mathematical theory of probabilistic parametric inference. In particular, in Section 2, the notions of probability, of (direct) probability distribution, of parametric family and of invariant family are introduced. In addition, some of the properties of probability distributions are briefly reviewed. In Section 3, the so-called inverse probability distributions are defined. It is demonstrated that the inverse probability distributions must be directly proportional to the appropriate direct probability distributions. The proportionality factors, called consistency factors, are determined in Section 4 on the grounds of invariance of parametric families of direct probability distributions under the action of Lie groups. In Section 5, the concept of rel-

(e-mail: Tomaz.Podobnik@ijs.si; Tomi.Zivko@ijs.si)

ative frequency and the concept of degree of belief are introduced that link the probability distributions to an external world of measurable phenomena. In this way, the mathematical theory becomes operational. Also in Section 5, as well as in Conclusions, a reconciliation between the Bayesian and the frequentist schools of parametric inference is advocated.

2. PROBABILITIES AND PROBABILITY DISTRIBUTIONS

2.1 Notation and general definitions

In this section, the notions of probability and of probability distribution are introduced, and some of the properties of probability distributions are briefly reviewed, with special attention being paid to conditional probability density functions. The purpose of refreshing these well known concepts is to avoid misunderstandings in subsequent sections where the properties of probability distributions are extensively invoked and the definition of the conditional probability distribution is extended.

Let Ω be a non-empty *universal set*, also called a *sample space*, whose elements are denoted by ω . A set Σ of subsets $A, B, C \dots$ of the sample space is called a σ -*algebra* (or σ -*field*) on Ω if Σ has Ω as a member, and is closed under complementation, $\bar{A} \in \Sigma$; $\forall A \in \Sigma$, and under countable union, $\sum_{i=1}^{\infty} A_i \in \Sigma$; $\forall A_1, A_2, \dots \in \Sigma$ (throughout the present discussion, $A + B$, AB and $A - B$ denote a union, an intersection and a relative complement of sets A and B , respectively, while $\bar{A} \equiv \Omega - A$). An ordered pair (Ω, Σ) consisting of a state space Ω and a σ -algebra Σ on Ω is called a *measurable space*.

EXAMPLE 1 (Borel algebra). Let Ω be \mathbb{R}^n . The *Borel σ -algebra* (or *Borel algebra*) \mathcal{B}^n on \mathbb{R}^n is the minimal σ -algebra containing a collection of open rectangles in \mathbb{R}^n . It is also said that the Borel algebra \mathcal{B}^n on \mathbb{R}^n is *generated* by all open rectangles in \mathbb{R}^n . Every set from a Borel algebra is called a *Borel set*.

DEFINITION 1 (Probability). Let P be a real-valued function on a σ -field Σ on a sample space Ω . We call P a probability measure (or simply a probability) if it is congruent with the following three

axioms due to Kolmogorov (1933):

$$\begin{aligned} (1) \quad & P(A) \geq 0 ; \quad \forall A \in \Sigma , \\ (2) \quad & P(\Omega) = 1 , \\ (3) \quad & P\left(\sum_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \end{aligned}$$

for all $A_i, A_j \in \Sigma$ that are mutually exclusive, i.e., $A_i A_j \neq i = \emptyset$. Then, the triple (Ω, Σ, P) is termed the probability space.

DEFINITION 2 (Random variable). Given a probability space (Ω, Σ, P) , let a function $X : \Omega \rightarrow \mathbb{R}$ be Σ -measurable: $A_{X \leq x} = \{\omega \in \Omega : X(\omega) \leq x\} \in \Sigma, \forall x \in \mathbb{R}$. Then, X is called a (real-valued) scalar random variable (or random variate), while x is called a realization of X .

DEFINITION 3 (Distribution function). Given a random variable X on a probability space (Ω, Σ, P) , the (cumulative) distribution function (cdf) $F_X(x)$ is a real-valued function on the state space \mathbb{R} to $[0, 1]$ such that $F_X(x) = P(A_{X \leq x})$.

Every cdf is a non-decreasing function with $F_X(-\infty) \equiv \lim_{x \rightarrow -\infty} F_X(x) = 0$ and

$$(4) \quad F_X(+\infty) \equiv \lim_{x \rightarrow +\infty} F_X(x) = 1 .$$

DEFINITION 4 (Continuous random variable). A random variable X is called continuous if its cdf $F_X(x)$ is absolutely continuous, i.e., if the cdf is expressible as an integral of a non-negative (Lebesgue) integrable function $f_X(x)$, called probability density function (pdf):

$$F_X(x) = \int_{-\infty}^x f_X(x') dx' .$$

The support of a continuous random variable X is a set, say V_X , of all x for which $f_X(x) > 0$.

Due to (4), a pdf is always normalized to unit area,

$$(5) \quad \int_{-\infty}^{+\infty} f_X(x') dx' = \int_{V_X} f_X(x') dx' = 1 .$$

Two pdf's correspond to the same cdf precisely if they differ only on a set of Lebesgue measure zero.

On the other hand, a cdf of a continuous random variable is differentiable almost everywhere on \mathbb{R} (Stein and Shakarachi (2005), § 3.2, Theorem 3.11, pp.130-131) such that the derivative can be used as a pdf.

DEFINITION 5. *Throughout the present discussion,*

$$(6) \quad f_X(x) \equiv \frac{d}{dx} F_X(x)$$

is assumed.

DEFINITION 6 (Probability distribution). *A function $Pr_X : \mathcal{B} \rightarrow [0, 1]$ called probability distribution is defined as the image measure of P by the random variable X , $Pr_X \equiv P \circ X^{-1}$, such that $Pr_X(S) = P[X^{-1}(S)]$, where $X^{-1}(S) \in \Sigma$ is the inverse image of a Borel set S under X . A probability distribution over a continuous random variable X is called a continuous probability distribution.*

From the properties of the underlying probability spaces it follows immediately that probability distributions for random variables also conform to the axioms (1-3) of probability. Therefore, a scalar random variate X on a probability space (Ω, Σ, P) generates another probability space $(\mathbb{R}, \mathcal{B}, Pr_X)$ with the Borel algebra $\mathcal{B} \equiv \mathcal{B}^1$ as underlying σ -algebra.

Let X and Y be continuous random variables defined on (Ω, Σ, P) , let there exist a function s on V_X such that $Y = s \circ X$ and $y = s(x)$, and let the function s be differentiable with non-vanishing derivative $s'(x)$ on the entire support V_X of X , such that $[s^{-1}(y)]' = [s'(x)]^{-1}$ exists for all $y = s(x)$ with $x \in V_X$. Then, due to the common probability space (Ω, Σ, P) underlying the spaces $(\mathbb{R}, \mathcal{B}, Pr_X)$ and $(\mathbb{R}, \mathcal{B}, Pr_Y)$,

$$\begin{array}{ccc} (\mathbb{R}, \mathcal{B}, Pr_X) & \xrightarrow{s} & (\mathbb{R}, \mathcal{B}, Pr_Y) \\ & \swarrow X \quad \searrow Y & \\ & (\Omega, \Sigma, P) & \end{array},$$

for all y for which $s^{-1}(y) \in V_X$ the cdf F_Y for Y can be expressed in terms of F_X as

$$(7) \quad F_Y(y) = \begin{cases} F_X(s^{-1}(y)) & ; [s^{-1}(y)]' > 0 \\ 1 - F_X(s^{-1}(y)) & ; [s^{-1}(y)]' < 0 \end{cases},$$

and the pdf for Y is related to the pdf for X as

$$(8) \quad f_Y(y) = \frac{d}{dy} F_Y(y) = f_X(s^{-1}(y)) |[s^{-1}(y)]'|.$$

The image of V_X under s is contained in V_Y , $s(V_X) \subseteq V_Y$, and the probability distribution $Pr_Y[V_Y - s(V_X)]$ for the relative complement of V_Y and $s(V_X)$ is zero.

The foregoing discussion about the probability distributions associated to scalar random variables is extended to multivariate random variables as follows.

DEFINITION 7 (Random vectors). *Given a probability space (Ω, Σ, P) , a vector function $\mathbf{X} = (X_1, \dots, X_n)$ is called a multivariate random variable (or random vector) if $A_{\mathbf{X} \leq \mathbf{x}} = \{\omega \in \Omega : X_1(\omega) \leq x_1, \dots, X_n(\omega) \leq x_n\} \in \Sigma, \forall \mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$. Every random vector gives rise to a cdf $F_{\mathbf{X}}(x_1, \dots, x_n)$ on the state space \mathbb{R}^n to $[0, 1]$ such that $F_{\mathbf{X}}(x_1, \dots, x_n) = P(A_{\mathbf{X} \leq \mathbf{x}})$, and to a joint probability distribution $Pr_{\mathbf{X}}(S)$ on the Borel algebra \mathcal{B}^n to $[0, 1]$, $Pr_{\mathbf{X}}(S) \equiv P[\mathbf{X}^{-1}(S)]$, $S \in \mathcal{B}^n$. Also, as for the scalar random variates, a random vector \mathbf{X} is called continuous if its cdf can be written as an integral of a pdf $f_{\mathbf{X}}(x_1, \dots, x_n)$,*

$$\begin{aligned} F_{\mathbf{X}}(x_1, \dots, x_n) &= \int_{U_{\mathbf{X} \leq \mathbf{x}}} f_{\mathbf{X}}(t_1, \dots, t_n) dt_1 \cdots dt_n \\ &= \int_{-\infty}^{x_1} dt_1 \cdots \int_{-\infty}^{x_n} dt_n f_{\mathbf{X}}(t_1, \dots, t_n), \end{aligned}$$

where $U_{\mathbf{X} \leq \mathbf{x}} \equiv \times_{i=1}^n (-\infty, x_i]$ is an infinite n -dimensional rectangle in the state space \mathbb{R}^n , while the transition from a n -dimensional integral to n iterated integrals is justified by Fubini's Theorem (see, for example, Bartle (1966), Chapter 10, pp.119-120).

Every (joint) probability distribution for a continuous n -vector \mathbf{X} can be expressed as an integral

$$Pr_{\mathbf{X}}(S) \equiv \int_S f_{\mathbf{X}}(\mathbf{x}) d^n \mathbf{x}; \quad \forall S \in \mathcal{B}^n.$$

Let \mathbf{X} and \mathbf{Y} be n -dimensional continuous random variables on a probability space (Ω, Σ, P) , let $f_{\mathbf{X}}(\mathbf{x})$ be a pdf for \mathbf{X} , and let \mathbf{s} be a differentiable function on $V_{\mathbf{X}}$ with non-vanishing Jacobian

$|\partial_{\mathbf{x}} \mathbf{s}(\mathbf{x})|$ such that $\mathbf{Y} = \mathbf{s} \circ \mathbf{X}$. Then, for all \mathbf{y} from the image of $V_{\mathbf{X}}$ under \mathbf{s} , the pdf for \mathbf{Y} reads:

$$(9) \quad f_{\mathbf{Y}}(\mathbf{y}) \equiv f_{\mathbf{X}}(\mathbf{s}^{-1}(\mathbf{y})) |\partial_{\mathbf{y}} \mathbf{s}^{-1}(\mathbf{y})|.$$

DEFINITION 8 (Marginal distributions). *Let a random vector \mathbf{X} be partitioned into a random n -vector \mathbf{Y} and a random m -vector \mathbf{Z} , $\mathbf{X} = (\mathbf{Y}, \mathbf{Z})$. Then $F_{\mathbf{X}}^{\mathbf{Y}}(\mathbf{y}) \equiv F_{\mathbf{X}}(\mathbf{y}, z_1 = \infty, \dots, z_m = \infty)$ and $F_{\mathbf{X}}^{\mathbf{Z}}(\mathbf{z}) \equiv F_{\mathbf{X}}(y_1 = \infty, \dots, y_n = \infty, \mathbf{z})$ are called the marginal cdf's for the components \mathbf{Y} and \mathbf{Z} of the partition (\mathbf{Y}, \mathbf{Z}) of \mathbf{X} , respectively. Also, pdf's*

$$f_{\mathbf{X}}^{\mathbf{Y}}(\mathbf{y}) \equiv \int_{\mathbb{R}^m} f_{\mathbf{X}}(\mathbf{y}, \mathbf{z}) d^m \mathbf{z}$$

and

$$f_{\mathbf{X}}^{\mathbf{Z}}(\mathbf{z}) \equiv \int_{\mathbb{R}^n} f_{\mathbf{X}}(\mathbf{y}, \mathbf{z}) d^n \mathbf{y}$$

are called the marginal pdf's for the components \mathbf{Y} and \mathbf{Z} of a partition of a continuous random vector \mathbf{X} , while the corresponding marginal probability distributions are denoted by $Pr_{\mathbf{X}}^{\mathbf{Y}}(U)$ and $Pr_{\mathbf{X}}^{\mathbf{Z}}(S)$, $U \in \mathcal{B}^n$ and $S \in \mathcal{B}^m$.

Usually, abbreviated notations may be used, e.g., $F_{\mathbf{X}}(\mathbf{y}) \equiv F_{\mathbf{X}}^{\mathbf{Y}}(\mathbf{y})$ and $f_{\mathbf{X}}(\mathbf{z}) \equiv f_{\mathbf{X}}^{\mathbf{Z}}(\mathbf{z})$. Since, however, in $F_{\mathbf{X}}(\mathbf{y})$ and in $f_{\mathbf{X}}(\mathbf{y})$ the arguments of the functions denote also the functions themselves, it should be noted that $F_{\mathbf{X}}(\mathbf{y})$ and $f_{\mathbf{X}}(\mathbf{y})$ are not necessarily the same functions as $F_{\mathbf{X}}(\mathbf{z})$ and $f_{\mathbf{X}}(\mathbf{z})$, respectively.

DEFINITION 9 (Conditional probability distributions). *Let (Ω, Σ, P) be a probability space and $\mathbf{X} = (\mathbf{Y}, \mathbf{Z}) : \Omega \rightarrow \mathbb{R}^n \times \mathbb{R}^m$ a Σ -measurable function that gives rise to a probability distribution $Pr_{\mathbf{X}} : \mathcal{B}^n \times \mathcal{B}^m \rightarrow [0, 1]$, let $(\mathbb{R}^n, \mathcal{B}^n, Pr_{\mathbf{X}}^{\mathbf{Y}})$ and $(\mathbb{R}^m, \mathcal{B}^m, Pr_{\mathbf{X}}^{\mathbf{Z}})$ be the spaces of the marginal probability distributions for the components \mathbf{Y} and \mathbf{Z} of the partition (\mathbf{Y}, \mathbf{Z}) of \mathbf{X} , and let $\mathbf{1}_{\mathbf{Y}^{-1}(U)}$, $U \in \mathcal{B}^n$, be the indicator function on Ω : $\mathbf{1}_{\mathbf{Y}^{-1}(U)}(\omega) = 1$ for $\omega \in \mathbf{Y}^{-1}(U)$ and 0 otherwise. Then, a function $\nu_{\mathbf{1}_{\mathbf{Y}^{-1}(U)}} : \Sigma' \rightarrow \mathbb{R}$, $\Sigma' \equiv \mathbf{Z}^{-1}(\mathcal{B}^m) \subset \Sigma$,*

$$\nu_{\mathbf{1}_{\mathbf{Y}^{-1}(U)}}[\mathbf{Z}^{-1}(S)] \equiv \int_{\mathbf{Z}^{-1}(S)} \mathbf{1}_{\mathbf{Y}^{-1}(U)}(\omega) dP(\omega),$$

$S \in \mathcal{B}^m$, is a finite measure on Σ' , and so is finite the image measure $\tilde{\nu}_{\mathbf{1}_{\mathbf{Y}^{-1}(U)}}$ of the measure $\nu_{\mathbf{1}_{\mathbf{Y}^{-1}(U)}}$ by \mathbf{Z} , $\tilde{\nu}_{\mathbf{1}_{\mathbf{Y}^{-1}(U)}} : \mathcal{B}^m \rightarrow \mathbb{R}$, $\tilde{\nu}_{\mathbf{1}_{\mathbf{Y}^{-1}(U)}} \equiv \nu_{\mathbf{1}_{\mathbf{Y}^{-1}(U)}} \circ$

\mathbf{Z}^{-1} . The function $Pr_{\mathbf{X}}^{\mathbf{Y}|\mathbf{Z}=\mathbf{z}}(U|\mathbf{z}) : \mathbb{R}^m \rightarrow \mathbb{R}$ called conditional probability distribution for \mathbf{Y} given the value $\mathbf{Z} = \mathbf{z}$, is then defined by the set of functional equations:

$$(10) \quad \tilde{\nu}_{\mathbf{1}_{\mathbf{Y}^{-1}(U)}}(S) = \int_S Pr_{\mathbf{X}}^{\mathbf{Y}|\mathbf{Z}=\mathbf{z}}(U|\mathbf{z})(\mathbf{z}) dPr_{\mathbf{X}}^{\mathbf{Z}}(\mathbf{z}),$$

while the corresponding conditional cdf is denoted by $F_{\mathbf{X}}^{\mathbf{Y}|\mathbf{Z}=\mathbf{z}}(\mathbf{y}|\mathbf{z})$.

The definition of $Pr_{\mathbf{X}}^{\mathbf{Y}|\mathbf{Z}=\mathbf{z}}(U|\mathbf{z})$ can be interpreted to say that the diagram

$$\begin{array}{ccc} & \mathbb{R} & \\ \uparrow \mathbf{1}_{\mathbf{Y}^{-1}(U)} & \swarrow Pr_{\mathbf{X}}^{\mathbf{Y}|\mathbf{Z}=\mathbf{z}}(U|\mathbf{z}) & \\ \Omega & \xleftarrow{\mathbf{Z}^{-1}} & \mathbb{R}^m \end{array}$$

is commutative in the average with respect to $Pr_{\mathbf{X}}^{\mathbf{Z}}$.

DEFINITION 10 (Conditional pdf). *Let $Pr_{\mathbf{X}}^{\mathbf{Y}|\mathbf{Z}=\mathbf{z}}(U|\mathbf{z})$ be a solution of (10). For continuous \mathbf{X} , the system of equations*

$$(11) \quad Pr_{\mathbf{X}}^{\mathbf{Y}|\mathbf{Z}=\mathbf{z}}(U|\mathbf{z}) = \int_U f_{\mathbf{X}}^{\mathbf{Y}|\mathbf{Z}=\mathbf{z}}(\mathbf{y}|\mathbf{z}) d^n \mathbf{y}$$

for all $U \in \mathcal{B}^n$, is the defining condition for the conditional pdf $f_{\mathbf{X}}^{\mathbf{Y}|\mathbf{Z}=\mathbf{z}}$ for \mathbf{Y} given $\mathbf{Z} = \mathbf{z}$.

For conditional cdf's and pdf's, abbreviated notations $F_{\mathbf{X}}(\mathbf{y}|\mathbf{z}) \equiv F_{\mathbf{X}}^{\mathbf{Y}|\mathbf{Z}=\mathbf{z}}(\mathbf{y}|\mathbf{z})$ and $f_{\mathbf{X}}(\mathbf{y}|\mathbf{z}) \equiv f_{\mathbf{X}}^{\mathbf{Y}|\mathbf{Z}=\mathbf{z}}(\mathbf{y}|\mathbf{z})$ may again be used.

PROPOSITION 1. *Let $f_{\mathbf{X}}(\mathbf{y}, \mathbf{z})$ be a joint pdf for a $(n+m)$ -dimensional random vector $\mathbf{X} = (\mathbf{Y}, \mathbf{Z})$ and let $f_{\mathbf{X}}(\mathbf{z})$ be the marginal pdf for \mathbf{Z} , supported on $V_{\mathbf{Z}}$. Then,*

$$(12) \quad f_{\mathbf{X}}(\mathbf{y}|\mathbf{z}) = \frac{f_{\mathbf{X}}(\mathbf{y}, \mathbf{z})}{f_{\mathbf{X}}(\mathbf{z})}$$

holds true uniquely on $(\mathbb{R}^n - U_0) \times (V_{\mathbf{Z}} - S_0)$, where $Pr_{\mathbf{X}}^{\mathbf{Z}}(S_0) = \nu_L(U_0) = 0$, $\nu_L(U_0) \equiv \int_{U_0} d^n \mathbf{y}$. It is said that $f_{\mathbf{X}}(\mathbf{y}|\mathbf{z})$ is determined uniquely $Pr_{\mathbf{X}}^{\mathbf{Z}}$ -almost everywhere on $V_{\mathbf{Z}}$ and ν_L -almost everywhere on \mathbb{R}^n .

REMARK 1. First, the reason for adopting an indirect definition of the conditional pdf's is that the

more direct formulations like, for example, the approach that is based on the L'Hôpital rule (see, for example, Rao (1993), § 1.4, pp. 13-14) and the axiomatization of Rényi (1955), do not lead to uniquely defined conditional pdf's. For a discussion on the resulting inconsistencies see Rao (1993), Chapters 3 and 4, pp. 63-121. Second, below, existence of a joint pdf $f_{\mathbf{X}}(\mathbf{y}, \mathbf{z})$ is not a necessary condition for existence of the corresponding conditional pdf's $f_{\mathbf{X}}(\mathbf{y}|\mathbf{z})$ and $f_{\mathbf{X}}(\mathbf{z}|\mathbf{y})$.

Let there exist a conditional pdf $f_{\mathbf{X}}(\mathbf{y}, \mathbf{z}|\mathbf{t})$, $\mathbf{X} = (\mathbf{Y}, \mathbf{Z}, \mathbf{T})$, and let the marginal distribution

$$f_{\mathbf{X}}(\mathbf{z}|\mathbf{t}) \equiv \int_{\mathbb{R}^{n_1}} f_{\mathbf{X}}(\mathbf{y}, \mathbf{z}|\mathbf{t}) d^{n_1} \mathbf{y}$$

be positive. Then, by an iterative application of Definition 10,

$$(13) \quad f_{\mathbf{X}}(\mathbf{y}|\mathbf{z}, \mathbf{t}) = \frac{f_{\mathbf{X}}(\mathbf{y}, \mathbf{z}|\mathbf{t})}{f_{\mathbf{X}}(\mathbf{z}|\mathbf{t})}.$$

The results of the following example are obtained by sequential applications of the *product rule* (13).

EXAMPLE 2. Let \mathbf{X} be partitioned into $(\mathbf{Y}, \mathbf{Z}, \mathbf{T}, \mathbf{W})$ and let there exist conditional pdf's $f_{\mathbf{X}}(\mathbf{y}, \mathbf{t}|\mathbf{z}, \mathbf{w})$ and $f_{\mathbf{X}}(\mathbf{y}, \mathbf{w}|\mathbf{z}, \mathbf{t})$. Then, in an analogy with (13), for $f_{\mathbf{X}}(\mathbf{t}|\mathbf{z}, \mathbf{w}), f_{\mathbf{X}}(\mathbf{w}|\mathbf{z}, \mathbf{t}) > 0$ there exists a conditional pdf $f_{\mathbf{X}}(\mathbf{y}|\mathbf{z}, \mathbf{t}, \mathbf{w})$ such that

$$f_{\mathbf{X}}(\mathbf{y}|\mathbf{z}, \mathbf{t}, \mathbf{w}) = \frac{f_{\mathbf{X}}(\mathbf{y}, \mathbf{t}|\mathbf{z}, \mathbf{w})}{f_{\mathbf{X}}(\mathbf{t}|\mathbf{z}, \mathbf{w})} = \frac{f_{\mathbf{X}}(\mathbf{y}, \mathbf{w}|\mathbf{z}, \mathbf{t})}{f_{\mathbf{X}}(\mathbf{w}|\mathbf{z}, \mathbf{t})}.$$

When, in addition, the marginal pdf's $f_{\mathbf{X}}(\mathbf{y}|\mathbf{z}, \mathbf{w})$ and $f_{\mathbf{X}}(\mathbf{y}|\mathbf{z}, \mathbf{t})$ are also non-vanishing, the joint pdf's $f_{\mathbf{X}}(\mathbf{y}, \mathbf{t}|\mathbf{z}, \mathbf{w})$ and $f_{\mathbf{X}}(\mathbf{y}, \mathbf{w}|\mathbf{z}, \mathbf{t})$ can be further decomposed as $f_{\mathbf{X}}(\mathbf{y}, \mathbf{t}|\mathbf{z}, \mathbf{w}) = f_{\mathbf{X}}(\mathbf{y}|\mathbf{z}, \mathbf{w}) f_{\mathbf{X}}(\mathbf{t}|\mathbf{y}, \mathbf{z}, \mathbf{w})$ and $f_{\mathbf{X}}(\mathbf{y}, \mathbf{w}|\mathbf{z}, \mathbf{t}) = f_{\mathbf{X}}(\mathbf{y}|\mathbf{z}, \mathbf{t}) f_{\mathbf{X}}(\mathbf{w}|\mathbf{y}, \mathbf{z}, \mathbf{t})$, such that

$$(14) \quad \begin{aligned} f_{\mathbf{X}}(\mathbf{y}|\mathbf{z}, \mathbf{t}, \mathbf{w}) &= \frac{f_{\mathbf{X}}(\mathbf{y}|\mathbf{z}, \mathbf{w}) f_{\mathbf{X}}(\mathbf{t}|\mathbf{y}, \mathbf{z}, \mathbf{w})}{f_{\mathbf{X}}(\mathbf{t}|\mathbf{z}, \mathbf{w})} \\ &= \frac{f_{\mathbf{X}}(\mathbf{y}|\mathbf{z}, \mathbf{t}) f_{\mathbf{X}}(\mathbf{w}|\mathbf{y}, \mathbf{z}, \mathbf{t})}{f_{\mathbf{X}}(\mathbf{w}|\mathbf{z}, \mathbf{t})}. \end{aligned}$$

In the same way,

$$(15) \quad \begin{aligned} f_{\mathbf{X}}(\mathbf{y}|\mathbf{t}, \mathbf{w}) &= \frac{f_{\mathbf{X}}(\mathbf{y}|\mathbf{w}) f_{\mathbf{X}}(\mathbf{t}|\mathbf{y}, \mathbf{w})}{f_{\mathbf{X}}(\mathbf{t}|\mathbf{w})} \\ &= \frac{f_{\mathbf{X}}(\mathbf{y}|\mathbf{t}) f_{\mathbf{X}}(\mathbf{w}|\mathbf{y}, \mathbf{t})}{f_{\mathbf{X}}(\mathbf{w}|\mathbf{t})} \end{aligned}$$

is obtained when \mathbf{X} is partitioned into $(\mathbf{Y}, \mathbf{T}, \mathbf{W})$.

EXAMPLE 3 (Transformations of conditional pdf's). Let $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ be a continuous $(n_1 + n_2)$ -dimensional random variable and $f_{\mathbf{X}}(\mathbf{x}_1|\mathbf{x}_2)$ be a conditional pdf. Let, in addition, $\mathbf{s} : V_{\mathbf{X}} \rightarrow \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ be a differentiable function such that $\mathbf{Y} \equiv (\mathbf{Y}_1, \mathbf{Y}_2) = \mathbf{s} \circ \mathbf{X} = (\mathbf{s}_1 \circ \mathbf{X}_1, \mathbf{s}_2 \circ \mathbf{X}_2)$ and that the Jacobian $|\partial_{\mathbf{x}} \mathbf{s}(\mathbf{x})| = |\partial_{\mathbf{x}_1} \mathbf{s}_1(\mathbf{x}_1)| |\partial_{\mathbf{x}_2} \mathbf{s}_2(\mathbf{x}_2)|$ does not vanish on the entire support $V_{\mathbf{X}}$ of \mathbf{X} . For $f_{\mathbf{X}}(\mathbf{s}_2^{-1}(\mathbf{y}_2)) > 0$, equations (8) and (9) applied to the conditional pdf $f_{\mathbf{Y}}(\mathbf{y}_1|\mathbf{y}_2) = f_{\mathbf{Y}}(\mathbf{y}_1, \mathbf{y}_2)/f_{\mathbf{Y}}(\mathbf{y}_2)$ then yield

$$(16) \quad \begin{aligned} f_{\mathbf{Y}}(\mathbf{y}_1|\mathbf{y}_2) &= \frac{f_{\mathbf{X}}(\mathbf{s}_1^{-1}(\mathbf{y}_1), \mathbf{s}_2^{-1}(\mathbf{y}_2))}{f_{\mathbf{X}}(\mathbf{s}_2^{-1}(\mathbf{y}_2))} |\partial_{\mathbf{y}_1} \mathbf{s}_1^{-1}(\mathbf{y}_1)| \\ &= f_{\mathbf{X}}(\mathbf{s}_1^{-1}(\mathbf{y}_1) | \mathbf{s}_2^{-1}(\mathbf{y}_2)) |\partial_{\mathbf{y}_1} \mathbf{s}_1^{-1}(\mathbf{y}_1)|. \end{aligned}$$

During the present discussion we allow for a possibility that a conditional pdf $f_{\mathbf{X}}(\mathbf{x}_1|\mathbf{x}_2)$ exists even when the corresponding joint pdf $f_{\mathbf{X}}(\mathbf{x}_1, \mathbf{x}_2)$ does not exist. When $f_{\mathbf{X}}(\mathbf{x}_1, \mathbf{x}_2)$ does not exist, however, the transformation (16) of the conditional pdf that is induced by the transformation of the random vector, ceased to be uniquely determined. In order to dismiss this ambiguity, the following definition, motivated by the preceding example, is adopted.

DEFINITION 11 (Transformations of conditional pdf's). Let there exist a conditional pdf $f_{\mathbf{X}}(\mathbf{x}_1|\mathbf{x}_2)$, $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ and $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$, and let a function $\mathbf{s} : (\mathbf{x}_1, \mathbf{x}_2) \rightarrow (\mathbf{s}_1(\mathbf{x}_1), \mathbf{s}_2(\mathbf{x}_2)) \equiv (\mathbf{y}_1, \mathbf{y}_2)$ be one-to-one and with non-vanishing Jacobian $|\partial_{\mathbf{x}_1} \mathbf{s}_1(\mathbf{x}_1)|$ on the entire support $V_{\mathbf{X}_1|\mathbf{x}_2}$ of $f_{\mathbf{X}}(\mathbf{x}_1|\mathbf{x}_2)$. Then, the conditional pdf $f_{\mathbf{Y}}(\mathbf{y}_1|\mathbf{y}_2)$, $\mathbf{Y} \equiv (\mathbf{s}_1 \circ \mathbf{X}_1, \mathbf{s}_2 \circ \mathbf{X}_2) \equiv (\mathbf{Y}_1, \mathbf{Y}_2)$, is defined as

$$(17) \quad f_{\mathbf{Y}}(\mathbf{y}_1|\mathbf{y}_2) \equiv f_{\mathbf{X}}(\mathbf{s}_1^{-1}(\mathbf{y}_1) | \mathbf{s}_2^{-1}(\mathbf{y}_2)) |\partial_{\mathbf{y}_1} \mathbf{s}_1^{-1}(\mathbf{y}_1)|,$$

where $\mathbf{s}_{1,2}^{-1}$ are the inverse functions of $\mathbf{s}_{1,2}$.

2.2 Parametric families of probability distributions

The term *parametric family* is used to describe a collection $I = \{Pr_{I, \boldsymbol{\theta}} : \boldsymbol{\theta} \in V_{\boldsymbol{\Theta}}\}$ of probability distributions that differ only in the value of a (possibly multi-dimensional) *parameter*, say $\boldsymbol{\Theta}$, i.e., a value $\boldsymbol{\theta}$ of $\boldsymbol{\Theta}$ determines a unique distribution within I . Therefore, a probability distribution for

a random n -vector \mathbf{X} , $Pr_{\mathbf{X}}(S_{\mathbf{X}})$, $S_{\mathbf{X}} \in \mathcal{B}^n$, that belongs to a particular parametric family I , is denoted by $Pr_{I,\theta}(S_{\mathbf{X}})$, whereas $F_{I,\theta}(\mathbf{x})$ stands for the corresponding cdf. Likewise, $f_{I,\theta}(\mathbf{x})$ denotes a unique pdf within a parametric family I of continuous probability distributions. A continuous probability distribution from a parametric family I is supported on a set $V_{\mathbf{X}} = V_{\mathbf{X}}(\theta)$ that may, in general, depend on the value θ of the parameter, while the range $V_{\Theta} \subseteq \mathbb{R}^m$ of admissible values of Θ is called a *parameter space*. In the present article, every considered parametric family is assumed to be *identifiable*: $Pr_{I,\theta_1} \neq Pr_{I,\theta_2}$ for $\theta_1 \neq \theta_2$, $\theta_{1,2} \in V_{\Theta}$.

EXAMPLE 4 (Reparameterization). Let $f_{I,\theta}(\mathbf{x})$ be a pdf for a random n -vector \mathbf{X} from a parametric family I and let \mathbf{s} be a one-to-one Borel function onto \mathbb{R}^n such that the Jacobian $|\partial_{\mathbf{x}}\mathbf{s}(\mathbf{x})|$ does not vanish anywhere on the support $V_{\mathbf{X}}(\theta)$ of \mathbf{X} . Then, according to (9), $f_{I',\theta}(\mathbf{y}) = f_{I,\theta}(\mathbf{s}^{-1}(\mathbf{y})) |\partial_{\mathbf{y}}\mathbf{s}^{-1}(\mathbf{y})|$, where $\mathbf{Y} \equiv \mathbf{s} \circ \mathbf{X}$, $\mathbf{y} \equiv \mathbf{s}(\mathbf{x})$ and \mathbf{s}^{-1} is the inverse function of \mathbf{s} , while indices I and I' indicate that probability distributions for \mathbf{X} and \mathbf{Y} in general belong to different (but isomorphic) parametric families. Let, in addition, $\bar{\mathbf{s}}$ be a one-to-one function on the parameter space $V_{\Theta} \subseteq \mathbb{R}^m$, such that $\lambda \equiv \bar{\mathbf{s}}(\theta)$. Then, $f_{I',\theta}(\mathbf{y})$ can be reparameterized as

$$(18) \quad \begin{aligned} f_{I',\lambda}(\mathbf{y}) &= f_{I',\bar{\mathbf{s}}^{-1}(\lambda)}(\mathbf{y}) \\ &= f_{I,\bar{\mathbf{s}}^{-1}(\lambda)}(\mathbf{s}^{-1}(\mathbf{y})) |\partial_{\mathbf{y}}\mathbf{s}^{-1}(\mathbf{y})|, \end{aligned}$$

where $\bar{\mathbf{s}}^{-1}$ is the inverse function of $\bar{\mathbf{s}}$.

There is a complete analogy between the transformation (18) and the transformations (16) and (17), such that every probability distribution from a parametric family can be regarded as a conditional distribution, i.e., as a distribution that is conditional upon the value of the parameter. Accordingly, we define $F_I(\mathbf{x}|\theta) \equiv F_{I,\theta}(\mathbf{x})$ and $Pr_I(S_{\Theta=\theta,\mathbf{X}}|\theta) \equiv Pr_{I,\theta}(S_{\mathbf{X}})$, $S_{\Theta=\theta,\mathbf{X}} \in \mathcal{B}^n$, and, for continuous \mathbf{X} ,

$$(19) \quad f_I(\mathbf{x}|\theta) \equiv f_{I,\theta}(\mathbf{x})$$

for all $\mathbf{x} \in \mathbb{R}^n$ and $\theta \in V_{\Theta} \subseteq \mathbb{R}^m$.

$F_I(\mathbf{x}|\theta)$, $Pr_I(S_{\Theta=\theta,\mathbf{X}}|\theta)$ and $f_I(\mathbf{x}|\theta)$ are underlain by a probability space $(\Omega_{\theta}, \Sigma_{\theta}, P)$ and by a $(m+n)$ -dimensional random variable $(\Theta, \mathbf{X}) : \Omega_{\theta} \rightarrow (\theta, \mathbb{R}^n)$ for all $\theta \in V_{\Theta}$, where every state

space (θ, \mathbb{R}^n) is a slice on $V_{\Theta} \times \mathbb{R}^n$ that corresponds to a particular value θ of a m -dimensional parameter Θ of the family I . The probability distributions $Pr_I(S_{\Theta=\theta,\mathbf{X}}|\theta)$ on Borel σ -algebras \mathcal{B}^n on such slices are called *direct probability distributions* and represent the first step towards a unified approach to random variables and parameters from parametric families. The second step is made in Section 3, where the notion of the inverse probability distribution is introduced.

REMARK 2. The results of Subsections 2.2 and 2.3 are independent of the preceding definitions. The only reason to define $F_I(\mathbf{x}|\theta)$ and $f_I(\mathbf{x}|\theta)$ already at this stage is to avoid unnecessary duplications in notation.

DEFINITION 12 (Independent random variables). When $f_I(\mathbf{x}|\mathbf{y}, \theta) = f_I(\mathbf{x}|\theta)$ and $f_I(\mathbf{y}|\mathbf{x}, \theta) = f_I(\mathbf{y}|\theta)$, the components \mathbf{X} and \mathbf{Y} of a continuous random vector (\mathbf{X}, \mathbf{Y}) are called independent random variables. When, in addition, $f_I(\mathbf{x}|\theta)$ and $f_I(\mathbf{y}|\theta)$ are the same functions, the variables \mathbf{X} and \mathbf{Y} are said to have identical probability distribution.

When the components \mathbf{X} and \mathbf{Y} of a random vector (\mathbf{X}, \mathbf{Y}) are independent random variables and the joint pdf $f_I(\mathbf{x}, \mathbf{y}|\theta)$ exists, the latter can be written as $f_I(\mathbf{x}, \mathbf{y}|\theta) = f_I(\mathbf{x}|\theta) f_I(\mathbf{y}|\theta)$.

DEFINITION 13 (Location and scale parameters). Suppose a cdf for a scalar random variable X from a parametric family I is of the form

$$(20) \quad F_I(x|\mu, \sigma) = \Phi\left(\frac{x - \mu}{\sigma}\right),$$

where μ is a realization of the first component of a two-dimensional parameter $\Theta = (\Theta_1, \Theta_2)$, whereas σ is a realization of its second component. Then, Θ_1 is called a location parameter and Θ_2 is called a scale parameter, while $V_{\Theta} = \mathbb{R} \times \mathbb{R}^+$.

When probability distributions from a location-scale family I are continuous, on the support $V_X(\mu, \sigma)$ of a distribution from the family the appropriate pdf is of the form

$$(21) \quad f_I(x|\mu, \sigma) = \frac{d}{dx} F_I(x|\mu, \sigma) = \frac{1}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right),$$

where $\phi(x) \equiv \Phi'(x)$. Except for $x = \mu$, every pdf (21) from a location-scale family can be written as a sum

$$f_I(x|\mu, \sigma) = c_+ f_{I^+}(x|\mu, \sigma) + c_- f_{I^-}(x|\mu, \sigma) ,$$

where

$$c_+ f_{I^+}(x|\mu, \sigma) \equiv \begin{cases} 0 & ; \frac{x-\mu}{\sigma} \leq 0 \\ f_I(x|\mu, \sigma) & ; \frac{x-\mu}{\sigma} > 0 \end{cases}$$

and

$$c_- f_{I^-}(x|\mu, \sigma) \equiv \begin{cases} f_I(x|\mu, \sigma) & ; \frac{x-\mu}{\sigma} < 0 \\ 0 & ; \frac{x-\mu}{\sigma} \geq 0 \end{cases} ,$$

while

$$c_+ \equiv \int_0^\infty \phi(u) du \quad \text{and} \quad c_- \equiv \int_{-\infty}^0 \phi(u) du .$$

For $c_\pm > 0$, there exist pdf's $f_{I^\pm}(x|\mu, \sigma) \equiv f_I(x|\mu, \sigma)/c_\pm$, which can be further reduced to

$$(22) \quad f_{I^\pm}(y|\lambda_1, \lambda_2 = 1) = \frac{1}{\lambda_2} e^{\frac{y-\lambda_1}{\lambda_2}} \phi\left(\pm e^{\frac{y-\lambda_1}{\lambda_2}}\right) \\ \equiv \tilde{\phi}_\pm(y - \lambda_1) ,$$

where $y \equiv \ln\{\pm(x - \mu)\}$ and $\lambda_1 \equiv \ln \sigma$. That is, every scale parameter for a location-scale family I^\pm is reducible to a location parameter for a parametric family $I^{\pm'}$.

2.3 Invariant families of probability distributions

Let $G = \{a, b, c, \dots\}$ be a group whose unit element is denoted by e and let \mathbf{l} be a function on $G \times \mathbb{R}^n$ to \mathbb{R}^n satisfying $\mathbf{l}(e, \mathbf{x}) = \mathbf{x}$, $\forall \mathbf{x} \in \mathbb{R}^n$ and $\mathbf{l}(a \circ b, \mathbf{x}) = \mathbf{l}[a, \mathbf{l}(b, \mathbf{x})]$, $\forall a, b \in G$ and $\forall \mathbf{x} \in \mathbb{R}^n$. Such a function specifies G acting on the left of \mathbb{R}^n and a group $\mathcal{G} = \{\mathbf{g}_a : a \in G\}$ of functions $\mathbf{g}_a : \mathbb{R}^n \rightarrow \mathbb{R}^n$, $\mathbf{g}_a(\mathbf{x}) \equiv \mathbf{l}(a, \mathbf{x})$. A composition of $\mathbf{g}_a, \mathbf{g}_b \in \mathcal{G}$ corresponds to the composition of $a, b \in G$, $\mathbf{g}_a[\mathbf{g}_b(\mathbf{x})] \equiv (\mathbf{g}_a \circ \mathbf{g}_b)(\mathbf{x}) = \mathbf{g}_{a \circ b}(\mathbf{x})$, \mathbf{g}_e is the unit element in \mathcal{G} and $\mathbf{g}_{a^{-1}} = \mathbf{g}_a^{-1}$, $\forall a \in G$ (see, for example, Eaton (1989), § 2.1, pp. 19-20).

DEFINITION 14 (Invariant family). *Let $F_I(\mathbf{x}|\boldsymbol{\theta})$ be a cdf from a parametric family I , let there exist a group G and a function $\mathbf{l} : G \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ specifying both an action of G on the left of the state space \mathbb{R}^n of the random n -vector \mathbf{X} and a group $\mathcal{G} = \{\mathbf{g}_a : a \in$*

$G\}$, $\mathbf{g}_a(\cdot) \equiv \mathbf{l}(a, \cdot)$, and let $\mathbf{Y} \equiv \mathbf{g}_a \circ \mathbf{X}$. In addition, let for every $\mathbf{g}_a \in \mathcal{G}$ and every $\boldsymbol{\theta} \in V_\Theta$ there exist a transformation $\bar{\mathbf{g}}_a : \boldsymbol{\theta} \rightarrow \bar{\mathbf{g}}_a(\boldsymbol{\theta}) \equiv \boldsymbol{\lambda}$, such that

$$(23) \quad F_{I'}(\mathbf{y}|\boldsymbol{\lambda}) = F_I(\mathbf{y}|\boldsymbol{\lambda}) .$$

where $\mathbf{y} \equiv \mathbf{g}_a(\mathbf{x})$. The family I is then said to be invariant under the group \mathcal{G} (or \mathcal{G} -invariant or invariant under the action of the group G).

Given a \mathcal{G} -invariant parametric family I , the set $\bar{\mathcal{G}} \equiv \{\bar{\mathbf{g}}_a : \mathbf{g}_a \in \mathcal{G}\}$ of the corresponding transformations on the parameter space is also a group (Ferguson (1967), § 4.1, Lemma 1, pp. 144-145), usually referred to as the *induced group* (Stuart et al. (1999), § 23.10, p. 300).

Let elements of a group G be defined by the values of n continuous real parameters (or *coordinates*), e.g., $a = \gamma(a_1, \dots, a_n)$ with γ being a function on (a subset of) \mathbb{R}^n to G . The coordinates are essential in the sense that the group elements cannot be distinguished by any number of coordinates smaller than the *dimension* n of the group G . Since, by definition, every group is closed under composition of its elements, $a \circ b = c \in G$, $\forall a, b \in G$, the coordinates of c are expressible as functions of the coordinates of a and b , $c_i = c_i(a_1, \dots, a_n; b_1, \dots, b_n)$, $i = 1, \dots, n$.

EXAMPLE 5 (One-dimensional groups). Coordinates of elements of a one-dimensional group G also form a group $\tilde{G} \subseteq \mathbb{R}$ with $a_1 \circ b_1 \equiv c_1(a_1, b_1)$ being the corresponding group operation in \tilde{G} . Therefore, since G and \tilde{G} are isomorphic, no generality is lost if $a = \gamma(a_1) = a_1$ is assumed.

When coordinates c_i of an element $c = a \circ b$ of a n -dimensional group G are smooth (i.e., C^∞) functions of the parameters of a and b , G is called *Lie group*.

EXAMPLE 6 (Invariance of location-scale families). $G = \mathbb{R} \times \mathbb{R}^+$ is a two-dimensional Lie group for the operations

$$(24) \quad a \circ b = (a_2 b_1 + a_1, a_2 b_2) .$$

Every location-scale family I (20) of continuous probability distributions is invariant under the group

$$(25) \quad \mathcal{G} = \{g_a : X \rightarrow a_2 X + a_1 ; (a_1, a_2) \in G\} ,$$

with

$$(26) \quad \bar{g}_a : (\Theta_1, \Theta_2) \longrightarrow (a_2\Theta_1 + a_1, a_2\Theta_2)$$

being the corresponding transformations from the induced group $\bar{\mathcal{G}}$. The family I is also invariant under two one-dimensional subgroups of the group \mathcal{G} : under the group $\mathcal{G}_\times = \{g_a : X \longrightarrow aX; a \in \mathbb{R}^+\}$, \mathbb{R}^+ is a one-dimensional Lie group for multiplication and $\{\bar{g}_a : (\Theta_1, \Theta_2) \longrightarrow (a\Theta_1, a\Theta_2); g_a \in \mathcal{G}_\times\}$ is the group induced by \mathcal{G}_\times , and under the group $\mathcal{G}_+ = \{g_a : X \longrightarrow a + X; a \in \mathbb{R}\}$, with \mathbb{R} being a one-dimensional Lie group for summation and with $\{\bar{g}_a : (\Theta_1, \Theta_2) \longrightarrow (a + \Theta_1, \Theta_2); g_a \in \mathcal{G}_+\}$ being the corresponding induced group.

Similarly, a family of continuous probability distributions for random vectors that consist of two independent scalar random variables X_1 and X_2 , both belonging to the same location-scale family I , is invariant under $\mathcal{G} = \{g_a : (X_1, X_2) \longrightarrow (a_2X_1 + a_1, a_2X_2 + a_1); (a_1, a_2) \in \mathbb{R} \times \mathbb{R}^+\}$, while the corresponding transformations from the induced group are again (26).

LEMMA 1. *Let G be a one-dimensional Lie group, let a function $l(a, x) : G \times \mathbb{R} \longrightarrow \mathbb{R}$ give rise to a group $\mathcal{G} = \{g_a : a \in G\}$ of transformations $g_a : \mathbb{R} \longrightarrow \mathbb{R}$, and let $l(a, x)$ be differentiable both in a and in x , $\forall a \in G$ and $\forall x \in \mathbb{R}$. Then, for all $x \in \mathbb{R}$ for which*

$$(27) \quad \partial_a l(a^{-1}, x)|_{a=e}$$

vanishes, all group transformations are trivial, i.e., $g_a(x) = x$ for all $g_a \in \mathcal{G}$.

Clearly, if (27) vanishes for all real x , then the action of the group G on the entire real axis is trivial: $g_a(x) = x$ for every $g_a \in \mathcal{G}$ and for all $x \in \mathbb{R}$.

LEMMA 2. *Suppose a probability distribution for a continuous scalar random variable X belongs to a family I of parametric distributions that is invariant under the action of a one-dimensional Lie group G . Let, in addition, the left actions $l(a, x)$ and $\bar{l}(a, \lambda)$ be differentiable in a , x and λ for all $a \in G$, $x \in V_X(\lambda)$ and $\lambda \in V_\Lambda(x)$, let the action of the group G not be identically trivial on the entire support $V_X(\lambda)$, and let the cdf for X , $F_I(x|\lambda)$, be differentiable in λ*

(differentiability in x is guaranteed by Definition 5). Then, the partial derivative

$$(28) \quad \partial_a \bar{l}(a^{-1}, \lambda)|_{a=e}$$

does not vanish anywhere on the space V_Λ of the (scalar) parameter Λ of the family I .

Furthermore, for a continuous scalar random variable X whose probability distribution belongs to a family I of parametric distributions that is invariant under the action of a group G , equation (7) reduces to

$$(29) \quad F_I(x|\lambda) = \begin{cases} F_I(l(a^{-1}, x)|\bar{l}(a^{-1}, \lambda)) & ; [g_a^{-1}(x)]' > 0 \\ 1 - F_I(l(a^{-1}, x)|\bar{l}(a^{-1}, \lambda)) & ; [g_a^{-1}(x)]' < 0 \end{cases},$$

$a \in G$. On the subspace $\tilde{V}_X \subseteq V_X(\lambda)$ with non-vanishing derivatives (27), derivatives $\partial_a \bar{l}(a^{-1}, \lambda)|_{a=e}$ are non-zero by Lemma 2. Then, for $x \in \tilde{V}_X$, differentiating (29) with respect to a and setting afterwards $a = e$ yields

$$(30) \quad \partial_x F_I(x|\lambda) \partial_\lambda H(x, \lambda) - \partial_\lambda F_I(x|\lambda) \partial_x H(x, \lambda) = 0,$$

where

$$(31) \quad H(x, \lambda) \equiv s(x) - \bar{s}(\lambda)$$

and

$$(32) \quad [s'(x)]^{-1} \equiv \partial_a h(a^{-1}, x)|_{a=e}$$

and

$$(33) \quad [\bar{s}'(\lambda)]^{-1} \equiv \partial_a \bar{h}(a^{-1}, \lambda)|_{a=e}.$$

LEMMA 3. *The cdf $F_I(x|\lambda)$ that solves the functional equation (30) is a differentiable function of a single variable $H(x, \lambda)$,*

$$(34) \quad F_I(x|\lambda) = \Phi[H(x, \lambda)].$$

Consequently, the cdf $F_I(x|\lambda)$ from a parametric family I that is invariant under the action of a one-dimensional Lie group can be written as

$$(35) \quad F_I(x|\lambda) = \Phi[s(x) - \bar{s}(\lambda)] = \Phi(y - \mu),$$

where $y \equiv s(x)$ and $\mu \equiv \bar{s}(\lambda)$ have been introduced. Then, by equation (7), the cdf for the continuous random variable $Y \equiv s \circ X$ is of the form

$$F_{I'}(y|\mu, \sigma = 1) = \begin{cases} \Phi(y - \mu) & ; [s^{-1}(y)]' > 0 \\ \tilde{\Phi}(y - \mu) & ; [s^{-1}(y)]' < 0 \end{cases},$$

where $\tilde{\Phi}(y - \mu) \equiv 1 - \Phi(y - \mu)$. That is, the probability distribution for the continuous random variable Y belongs to a location-scale family I' with $\sigma = 1$ (recall equation (20)), and the above reasoning can be summarized as

PROPOSITION 2. *Let X be a continuous scalar random variable whose probability distribution belongs to a \mathcal{G} -invariant parametric family I , where $\mathcal{G} = \{g_a : a \in G\}$ is underlain by a one-dimensional Lie group G . Let, in addition, $g_a(x)$ be differentiable for all $x \in \mathbb{R}$ and let the cdf $F_I(x|\lambda)$ for X be differentiable in λ . Then, on the subspace $\tilde{V}_X \subseteq V_X(\lambda)$ with non-vanishing derivatives (27), X is reducible by a one-to-one transformation s (32) to a continuous random variable $Y \equiv s \circ X$ whose probability distribution is from a location-scale family (20) with $\sigma = 1$ and $\mu \equiv \bar{s}(\lambda)$, where \bar{s} is defined via (33).*

REMARK 3. In the sequel (Proposition 4) we shall further demonstrate that for realizations $x \in V_X(\lambda) - \tilde{V}_X$ with vanishing derivative (27), a pdf cannot be assigned to the inferred parameter of the family I .

Let a continuous random variable \mathbf{X} with a pdf $f_I(\mathbf{x}|\boldsymbol{\theta})$ belong to a parametric family I that is invariant under a group \mathcal{G} of differentiable transformations \mathbf{g}_a with non-vanishing Jacobian $|\partial_{\mathbf{x}} \mathbf{g}_a(\mathbf{x})|$ on the entire support $V_{\mathbf{X}}(\boldsymbol{\theta})$ of the distribution for \mathbf{X} . Then, equation (9) applies which, when combined with the definition (23) of invariance of a family I , yields

$$f_I(\mathbf{y}|\boldsymbol{\lambda}) = f_I(\mathbf{g}_a^{-1}(\mathbf{y})|\bar{\mathbf{g}}_a^{-1}(\boldsymbol{\lambda})) |\partial_{\mathbf{y}} \mathbf{g}_a^{-1}(\mathbf{y})|$$

for all $\mathbf{y} \equiv \mathbf{g}_a(\mathbf{x})$ such that $\mathbf{x} \in V_{\mathbf{X}}(\boldsymbol{\theta})$, where $\boldsymbol{\lambda} \equiv \bar{\mathbf{g}}_a(\boldsymbol{\theta})$ and $\bar{\mathbf{g}}_a \in \bar{\mathcal{G}}$.

3. INVERSE PROBABILITY DISTRIBUTIONS

DEFINITION 15 (Inverse probability distributions). *Suppose there exist probability spaces $(\Omega_{\boldsymbol{\theta}}, \Sigma_{\boldsymbol{\theta}}, P)$, $\Omega_{\boldsymbol{\theta}} \subset \Omega$, for all $\boldsymbol{\theta} \in V_{\boldsymbol{\theta}}$ and a random variable $(\boldsymbol{\Theta}, \mathbf{X}) : \Omega_{\boldsymbol{\theta}} \rightarrow (\boldsymbol{\theta}, \mathbb{R}^n)$ that together lead to the parametric family I of continuous direct probability distributions $\text{Pr}_I(S_{\boldsymbol{\theta}=\boldsymbol{\theta}, \mathbf{X}}|\boldsymbol{\theta})$, $S_{\boldsymbol{\theta}=\boldsymbol{\theta}, \mathbf{X}} \in \mathcal{B}^n$ whose pdf's are denoted by $f_I(\mathbf{x}|\boldsymbol{\theta})$. Let, in addition, for some of those realizations \mathbf{x} of \mathbf{X} for which*

$$(36) \quad \int_{V_{\boldsymbol{\theta}}} f_I(\mathbf{x}|\boldsymbol{\theta}) d^m \boldsymbol{\theta} > 0 ,$$

there exist also probability spaces $(\Omega_{\mathbf{x}}, \Sigma_{\mathbf{x}}, P)$, $\Omega_{\mathbf{x}} \subset \Omega$, such that the function $(\boldsymbol{\Theta}, \mathbf{X}) : \Omega_{\mathbf{x}} \rightarrow (V_{\boldsymbol{\theta}}, \mathbf{x})$ is $\Sigma_{\mathbf{x}}$ -measurable (i.e., $A_{\boldsymbol{\theta} \leq \boldsymbol{\theta}} \equiv \{\omega \in \Omega_{\mathbf{x}} : \boldsymbol{\theta} \leq \boldsymbol{\theta}\} \in \Sigma_{\mathbf{x}}$ for all $\boldsymbol{\theta} \in V_{\boldsymbol{\theta}}$) and thus a random variable also on $(\Omega_{\mathbf{x}}, \Sigma_{\mathbf{x}}, P)$. Then, the probability distributions, resulting from the probability spaces $(\Omega_{\mathbf{x}}, \Sigma_{\mathbf{x}}, P)$ and from the corresponding random variable $(\boldsymbol{\Theta}, \mathbf{X})$, are called inverse probability distributions. The cdf's and the pdf's that correspond to the inverse probability distributions are denoted by $F_I(\boldsymbol{\theta}|\mathbf{x})$ and $f_I(\boldsymbol{\theta}|\mathbf{x})$, respectively.

Likewise, let $(\boldsymbol{\Theta}, \mathbf{X})$ be further partitioned into $(\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2, \mathbf{X})$ and let for some of those realizations $\boldsymbol{\theta}_1$ and \mathbf{x} for which

$$(37) \quad \int_{V_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2}} f_I(\mathbf{x}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) d^{m_1} \boldsymbol{\theta}_1 > 0 ,$$

there exist probability spaces $(\Omega_{\boldsymbol{\theta}_2, \mathbf{x}}, \Sigma_{\boldsymbol{\theta}_2, \mathbf{x}}, P)$ such that the function $(\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2, \mathbf{X}) : \Omega_{\boldsymbol{\theta}_2, \mathbf{x}} \rightarrow (V_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2}, \boldsymbol{\theta}_2, \mathbf{x})$ is $\Sigma_{\boldsymbol{\theta}_2, \mathbf{x}}$ -measurable, $A_{\boldsymbol{\theta}_1 \leq \boldsymbol{\theta}_1} \equiv \{\omega \in \Omega_{\boldsymbol{\theta}_2, \mathbf{x}} : \boldsymbol{\theta}_1 \leq \boldsymbol{\theta}_1\} \in \Sigma_{\boldsymbol{\theta}_2, \mathbf{x}}$ for all $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \in V_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2}$. Then, the cdf's and the pdf's that correspond to the resulting inverse probability distributions are denoted by $F_I(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2, \mathbf{x})$ and $f_I(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2, \mathbf{x})$, respectively.

REMARK 4. The integrals (36) and (37) need not be finite. The reasons for requiring the two integrals to be strictly positive will become apparent within the context of Proposition 3, below.

Apart from the direct and the inverse probability distributions, their mixtures may also exist. For example, $F_I(\boldsymbol{\theta}, \mathbf{x}_1|\mathbf{x}_2)$, $F_I(\boldsymbol{\theta}_1, \mathbf{x}_1|\boldsymbol{\theta}_2, \mathbf{x}_2)$, $f_I(\boldsymbol{\theta}, \mathbf{x}_1|\mathbf{x}_2)$ and $f_I(\boldsymbol{\theta}_1, \mathbf{x}_1|\boldsymbol{\theta}_2, \mathbf{x}_2)$ are the cdf's and the pdf's of two of the distributions that are neither purely direct nor purely inverse.

From a mathematical perspective, the direct and the inverse probability distributions, as well as their mixtures, share identical properties, some of which were discussed in Section 2.1. The following three rules that apply to inverse probability distributions are obtained by invoking the equivalence between the two types of distributions.

RULE 1 (Parameter transformation). *Let $f_I(\boldsymbol{\theta}|\mathbf{x})$ be a pdf of an inverse probability distribution and let $(\bar{\mathbf{s}}, \mathbf{s}) : (\boldsymbol{\Theta}, \mathbf{X}) \rightarrow (\bar{\mathbf{s}} \circ \boldsymbol{\Theta}, \mathbf{s} \circ \mathbf{X}) \equiv (\boldsymbol{\Lambda}, \mathbf{Y})$ be a differentiable transformation with a non-vanishing Jacobian on the entire support of $f_I(\boldsymbol{\theta}|\mathbf{x})$. Then,*

an inverse pdf $f_{I'}(\lambda|\mathbf{y})$ also exists and is related to $f_I(\theta|\mathbf{x})$ as

$$(38) \quad f_{I'}(\lambda|\mathbf{y}) = f_I(\bar{\mathbf{s}}^{-1}(\lambda)|\mathbf{s}^{-1}(\mathbf{y})) |\partial_{\lambda}\bar{\mathbf{s}}^{-1}(\lambda)|.$$

Similarly, when there exist an inverse pdf $f_I(\theta_1|\theta_2, \mathbf{x})$ and a differentiable transformation $(\bar{\mathbf{s}}_1, \bar{\mathbf{s}}_2, \mathbf{s}) : (\Theta_1, \Theta_2, \mathbf{X}) \longrightarrow (\bar{\mathbf{s}}_1 \circ \Theta, \bar{\mathbf{s}}_2 \circ \Theta, \mathbf{s} \circ \mathbf{X}) \equiv (\Lambda_1, \Lambda_2, \mathbf{Y})$ with a non-vanishing Jacobian on the support of $f_I(\theta_1|\theta_2, \mathbf{x})$, there exists a pdf $f_{I'}(\lambda_1|\lambda_2, \mathbf{y})$ such that

$$(39) \quad f_{I'}(\lambda_1|\lambda_2, \mathbf{y}) = f_I(\bar{\mathbf{s}}_1^{-1}(\lambda_1)|\bar{\mathbf{s}}_2^{-1}(\lambda_2), \mathbf{s}^{-1}(\mathbf{y})) |\partial_{\lambda_1}\bar{\mathbf{s}}_1^{-1}(\lambda_1)|.$$

Proof. If $f_I(\theta, \mathbf{x})$ exists, equation (38) follows from (16) by substitutions $\mathbf{X}_1 \rightarrow \Theta$, $\mathbf{X}_2 \rightarrow \mathbf{X}$, $\mathbf{Y}_1 \rightarrow \Lambda$, $\mathbf{Y}_2 \rightarrow \mathbf{Y}$, $\mathbf{s}_1 \rightarrow \bar{\mathbf{s}}$ and $\mathbf{s}_2 \rightarrow \mathbf{s}$. Similarly, if $f_I(\theta_1, \theta_2, \mathbf{x})$ exists, (39) is deduced from (16) by substitutions $\mathbf{X}_1 \rightarrow \Theta_1$, $\mathbf{X}_2 \rightarrow (\Theta_2, \mathbf{X})$, $\mathbf{Y}_1 \rightarrow \Lambda_1$, $\mathbf{Y}_2 \rightarrow (\Lambda_2, \mathbf{Y})$, $\mathbf{s}_1 \rightarrow \bar{\mathbf{s}}_1$ and $\mathbf{s}_2 \rightarrow (\bar{\mathbf{s}}_2, \mathbf{s})$. If, on the other hand, the joint pdf's $f_I(\theta, \mathbf{x})$ and $f_I(\theta_1, \theta_2, \mathbf{x})$ do not exist, equations (38) and (39) are definitions for $f_{I'}(\lambda|\mathbf{y})$ and $f_{I'}(\lambda_1|\lambda_2, \mathbf{y})$, respectively, in the same way as $f_{\mathbf{Y}}(\mathbf{y}_1|\mathbf{y}_2)$ was defined by (17). \square

RULE 2 (Product rule). *Let there exist an inverse pdf $f_I(\theta_1, \theta_2|\mathbf{x})$ and the corresponding marginal pdf*

$$f_I(\theta_2|\mathbf{x}) \equiv \int_{V_{\Theta_1}} f_I(\theta_1, \theta_2|\mathbf{x}) d^{m_1}\theta_1.$$

Then, for all θ_2 and \mathbf{x} for which $f_I(\theta_2|\mathbf{x}) > 0$,

$$(40) \quad f_I(\theta_1|\theta_2, \mathbf{x}) = \frac{f_I(\theta_1, \theta_2|\mathbf{x})}{f_I(\theta_2|\mathbf{x})}$$

holds uniquely (Lebesgue measure) ν_L -almost everywhere on \mathbb{R}^{m_1} .

Proof. The product rule (40) follows immediately from (13) by making substitutions $\mathbf{x}_1 \rightarrow \theta_1$, $\mathbf{x}_2 \rightarrow \theta_2$ and $\mathbf{x}_3 \rightarrow \mathbf{x}$. \square

RULE 3 (Bayes' Theorem). *Let a random vector be partitioned into $(\Theta_1, \Theta_2, \mathbf{X}_1, \mathbf{X}_2)$, let there exist pdf's $f_I(\theta_1, \mathbf{x}_1|\theta_2, \mathbf{x}_2)$ and $f_I(\theta_1, \mathbf{x}_2|\theta_2, \mathbf{x}_1)$, let marginal pdf's $f_I(\mathbf{x}_1|\theta_2, \mathbf{x}_2)$, $f_I(\mathbf{x}_2|\theta_2, \mathbf{x}_1)$, $f_I(\theta_1|\theta_2, \mathbf{x}_2)$ and $f_I(\theta_1|\theta_2, \mathbf{x}_1)$ be non-vanishing, and let the components \mathbf{X}_1 and \mathbf{X}_2 of the partition be*

independent random variables: $f_I(\mathbf{x}_1|\theta_1, \theta_2, \mathbf{x}_2) = f_I(\mathbf{x}_1|\theta_1, \theta_2)$ and $f_I(\mathbf{x}_2|\theta_1, \theta_2, \mathbf{x}_1) = f_I(\mathbf{x}_2|\theta_1, \theta_2)$. Then, there exists a conditional pdf $f_I(\theta_1|\theta_2, \mathbf{x}_1, \mathbf{x}_2)$ such that

$$(41) \quad f_I(\theta_1|\theta_2, \mathbf{x}_1, \mathbf{x}_2) = \frac{f_I(\theta_1|\theta_2, \mathbf{x}_2) f_I(\mathbf{x}_1|\theta_1, \theta_2)}{f_I(\mathbf{x}_1|\theta_2, \mathbf{x}_2)} = \frac{f_I(\theta_1|\theta_2, \mathbf{x}_1) f_I(\mathbf{x}_2|\theta_1, \theta_2)}{f_I(\mathbf{x}_2|\theta_2, \mathbf{x}_1)}.$$

If, on the other hand, a random vector is partitioned into $(\Theta, \mathbf{X}_1, \mathbf{X}_2)$,

$$(42) \quad f_I(\theta|\mathbf{x}_1, \mathbf{x}_2) = \frac{f_I(\theta|\mathbf{x}_2) f_I(\mathbf{x}_1|\theta)}{f_I(\mathbf{x}_1|\mathbf{x}_2)} = \frac{f_I(\theta|\mathbf{x}_1) f_I(\mathbf{x}_2|\theta)}{f_I(\mathbf{x}_2|\mathbf{x}_1)}$$

holds true under analogous conditions.

Proof. Equation (41) follows from (14) by making substitutions $\mathbf{y} \rightarrow \theta_1$, $\mathbf{z} \rightarrow \theta_2$, $\mathbf{t} \rightarrow \mathbf{x}_1$ and $\mathbf{w} \rightarrow \mathbf{x}_2$, whereas (42) is obtained from (15) by substitutions $\mathbf{y} \rightarrow \theta$, $\mathbf{t} \rightarrow \mathbf{x}_1$ and $\mathbf{w} \rightarrow \mathbf{x}_2$. \square

Equations (41) and (42) are also referred to as Bayes' Theorem (Bayes (1763); Laplace (1774)) or the principle of inverse probability (Jeffreys (1961), §1.22, p.28), written in terms of pdf's. In the equations, $f_I(\theta_1|\theta_2, \mathbf{x}_1, \mathbf{x}_2)$ and $f_I(\theta|\mathbf{x}_1, \mathbf{x}_2)$ are called the *posterior pdf's*, $f_I(\theta_1|\theta_2, \mathbf{x}_{1,2})$ and $f_I(\theta|\mathbf{x}_{1,2})$ are the so-called *prior pdf's*, $f_I(\mathbf{x}_{1,2}|\theta_1, \theta_2)$ and $f_I(\mathbf{x}_{1,2}|\theta)$ are the *likelihood densities*, while $f_I(\mathbf{x}_{1,2}|\theta_2, \mathbf{x}_{2,1})$ and $f_I(\mathbf{x}_{1,2}|\mathbf{x}_{2,1})$ are the *predictive pdf's*. While the predictive pdf's are determined by the normalization condition on the posterior pdf's, e.g.,

$$f_I(\mathbf{x}_{1,2}|\theta_2, \mathbf{x}_{2,1}) = \int_{V_{\Theta_1}} f_I(\theta_1|\theta_2, \mathbf{x}_{2,1}) f_I(\mathbf{x}_{1,2}|\theta_1, \theta_2) d^{m_1}\theta_1,$$

the general form of the prior pdf's $f_I(\theta_1|\theta_2, \mathbf{x}_{1,2})$ and $f_I(\theta|\mathbf{x}_{1,2})$ is prescribed by the following Proposition.

PROPOSITION 3. *Suppose that conditions for Bayes' Theorem (41) are fulfilled: a random vector is partitioned into $(\Theta_1, \Theta_2, \mathbf{X}_1, \mathbf{X}_2)$, there exist conditional pdf's $f_I(\theta_1, \mathbf{x}_1|\theta_2, \mathbf{x}_2)$ and $f_I(\theta_1, \mathbf{x}_2|\theta_2, \mathbf{x}_1)$,*

the marginal pdf's $f_I(\mathbf{x}_1|\boldsymbol{\theta}_2, \mathbf{x}_2)$, $f_I(\mathbf{x}_2|\boldsymbol{\theta}_2, \mathbf{x}_1)$, $f_I(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2, \mathbf{x}_2)$ and $f_I(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2, \mathbf{x}_1)$ are positive, and the components \mathbf{X}_1 and \mathbf{X}_2 of the partition are independent random variables with identical probability distribution. In addition, let $V_{\boldsymbol{\Theta}} = (V_{\boldsymbol{\Theta}_1}, V_{\boldsymbol{\Theta}_2})$ stand for the space of the parameter $\boldsymbol{\Theta} = (\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2)$ and let $\tilde{V}_{\boldsymbol{\Theta}_1}(\mathbf{x}_{1,2}, \boldsymbol{\theta}_2) \equiv \{\boldsymbol{\theta}_1 \in V_{\boldsymbol{\Theta}_1} : f_I(\mathbf{x}_{1,2}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) > 0\}$. Then, for $\boldsymbol{\theta}_1 \in \tilde{V}_{\boldsymbol{\Theta}_1}(\mathbf{x}_{1,2}, \boldsymbol{\theta}_2)$,

$$(43) \quad f_I(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2, \mathbf{x}_{1,2}) = \frac{\zeta_{I, \boldsymbol{\Theta}_1|\boldsymbol{\theta}_2}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\eta_{I, \boldsymbol{\Theta}_1|\boldsymbol{\theta}_2}(\mathbf{x}_{1,2}, \boldsymbol{\theta}_2)} f_I(\mathbf{x}_{1,2}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$$

is the most general form of the pdf's $f_I(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2, \mathbf{x}_{1,2})$. Similarly, when a random vector is partitioned into $(\boldsymbol{\Theta}, \mathbf{X}_1, \mathbf{X}_2)$, the conditions for Bayes' Theorem (42) are fulfilled and $\boldsymbol{\theta} \in \tilde{V}_{\boldsymbol{\Theta}}(\mathbf{x}_{1,2})$, $\tilde{V}_{\boldsymbol{\Theta}}(\mathbf{x}_{1,2}) \equiv \{\boldsymbol{\theta} \in V_{\boldsymbol{\Theta}} : f_I(\mathbf{x}_{1,2}|\boldsymbol{\theta}) > 0\}$,

$$(44) \quad f_I(\boldsymbol{\theta}|\mathbf{x}_{1,2}) = \frac{\zeta_{I, \boldsymbol{\Theta}}(\boldsymbol{\theta})}{\eta_{I, \boldsymbol{\Theta}}(\mathbf{x}_{1,2})} f_I(\mathbf{x}_{1,2}|\boldsymbol{\theta})$$

is the most general form of the pdf's $f_I(\boldsymbol{\theta}|\mathbf{x}_{1,2})$. The functions $\zeta_{I, \boldsymbol{\Theta}_1|\boldsymbol{\theta}_2}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ and $\zeta_{I, \boldsymbol{\Theta}}(\boldsymbol{\theta})$ in equations (43) and (44) are called the consistency factors.

Domains of $f_I(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2, \mathbf{x}_{1,2})$ and $f_I(\boldsymbol{\theta}|\mathbf{x}_{1,2})$ are extended beyond the supports $V_{\mathbf{X}_1}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = V_{\mathbf{X}_2}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ on which $f_I(\mathbf{x}_{1,2}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ and $f_I(\mathbf{x}_{1,2}|\boldsymbol{\theta})$ are positive by defining

$$f_I(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2, \mathbf{x}_{1,2}) \equiv \frac{\zeta_{I, \boldsymbol{\Theta}_1|\boldsymbol{\theta}_2}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\eta_{I, \boldsymbol{\Theta}_1|\boldsymbol{\theta}_2}(\mathbf{x}_{1,2}, \boldsymbol{\theta}_2)} f_I(\mathbf{x}_{1,2}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$$

for all $\mathbf{x}_{1,2} \notin V_{\mathbf{X}_{1,2}}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ and

$$f_I(\boldsymbol{\theta}|\mathbf{x}_{1,2}) \equiv \frac{\zeta_{I, \boldsymbol{\Theta}}(\boldsymbol{\theta})}{\eta_{I, \boldsymbol{\Theta}}(\mathbf{x}_{1,2})} f_I(\mathbf{x}_{1,2}|\boldsymbol{\theta})$$

for all $\mathbf{x}_{1,2} \notin V_{\mathbf{X}_{1,2}}(\boldsymbol{\theta})$. For the sake of symmetry between the direct and the inverse probability distributions, the domains of the inverse pdf's may be extended even further by defining $f_I(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2, \mathbf{x}_{1,2}) \equiv 0$ for $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \notin V_{\boldsymbol{\Theta}_1, \boldsymbol{\theta}_2}$ and $f_I(\boldsymbol{\theta}|\mathbf{x}_{1,2}) \equiv 0$ for $\boldsymbol{\theta} \notin V_{\boldsymbol{\Theta}}$. In this way, the inverse probability distribution spaces $(V_{\mathbf{x}}, \Sigma_{\mathbf{x}}, Pr)$ and $(V_{\boldsymbol{\theta}_2, \mathbf{x}}, \Sigma_{\boldsymbol{\theta}_2, \mathbf{x}}, Pr)$ are also extended to $(\mathbb{R}^m, \mathcal{B}^m, Pr)$ and $(\mathbb{R}^{m_1}, \mathcal{B}^{m_1}, Pr)$, respectively. Then, the *normalization factors* $\eta_{I, \boldsymbol{\Theta}_1|\boldsymbol{\theta}_2}(\mathbf{x}_{1,2}, \boldsymbol{\theta}_2)$ and $\eta_{I, \boldsymbol{\Theta}}(\mathbf{x}_{1,2})$ are determined by invoking normalization of the pdf's

$f_I(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2, \mathbf{x}_{1,2})$ and $f_I(\boldsymbol{\theta}|\mathbf{x}_{1,2})$:

$$\eta_{I, \boldsymbol{\Theta}_1|\boldsymbol{\theta}_2}(\mathbf{x}_{1,2}, \boldsymbol{\theta}_2) = \int_{\mathbb{R}^{m_1}} \zeta_{I, \boldsymbol{\Theta}_1|\boldsymbol{\theta}_2}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) f_I(\mathbf{x}_{1,2}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) d^{m_1} \boldsymbol{\theta}_1$$

and

$$\eta_{I, \boldsymbol{\Theta}}(\mathbf{x}_{1,2}) = \int_{\mathbb{R}^m} \zeta_{I, \boldsymbol{\Theta}}(\boldsymbol{\theta}) f_I(\mathbf{x}_{1,2}|\boldsymbol{\theta}) d^m \boldsymbol{\theta}.$$

Non-vanishing integrals (36) and (37) thus represent necessary conditions for normalizability (5) of the inverse pdf's $f_I(\boldsymbol{\theta}|\mathbf{x}_{1,2})$ and $f_I(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2, \mathbf{x}_{1,2})$.

For discrete random variables \mathbf{X}_1 and \mathbf{X}_2 , the appropriate forms of the pdf's $f_I(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2, \mathbf{x}_{1,2})$ and $f_I(\boldsymbol{\theta}|\mathbf{x}_{1,2})$ are obtained by replacing the likelihood densities $f_I(\mathbf{x}_{1,2}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ and $f_I(\mathbf{x}_{1,2}|\boldsymbol{\theta})$ in (43) and (44) with the *probability mass functions* $p_I(\mathbf{x}_{1,2}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ and $p_I(\mathbf{x}_{1,2}|\boldsymbol{\theta})$ that coincide with probability distributions for the points $\mathbf{X}_{1,2} = \mathbf{x}_{1,2}$ of a state space \mathbb{R}^n of the variables \mathbf{X}_1 and \mathbf{X}_2 , given the realizations $(\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2) = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ and $\boldsymbol{\Theta} = \boldsymbol{\theta}$ of the corresponding parameters.

REMARK 5. In equations (43) and (44), the pdf's $f_I(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2, \mathbf{x}_{1,2})$ and $f_I(\boldsymbol{\theta}|\mathbf{x}_{1,2})$ are directly proportional to the pdf's $f_I(\mathbf{x}_{1,2}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ and $f_I(\mathbf{x}_{1,2}|\boldsymbol{\theta})$ of the corresponding direct probability distributions. This is very similar to equations (41) and (42) of Bayes' Theorem with the posterior pdf's $f_I(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2, \mathbf{x}_1, \mathbf{x}_2)$ and $f_I(\boldsymbol{\theta}|\mathbf{x}_1, \mathbf{x}_2)$ being proportional to the likelihood densities $f_I(\mathbf{x}_{1,2}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ and $f_I(\mathbf{x}_{1,2}|\boldsymbol{\theta})$. But there is also a fundamental difference between the equations of Bayes' Theorem and those of Proposition 3: while the proportionality coefficients $f_I(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2, \mathbf{x}_{1,2})$ and $f_I(\boldsymbol{\theta}|\mathbf{x}_{1,2})$ between the posterior pdf's and the likelihood densities in Bayes' Theorem are the prior pdf's, the consistency factors $\zeta_{I, \boldsymbol{\Theta}_1|\boldsymbol{\theta}_2}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ and $\zeta_{I, \boldsymbol{\Theta}}(\boldsymbol{\theta})$ that are proportionality coefficients between the inverse and the direct pdf's in (43) and (44) need not be congruent with all the properties of probability density functions and should therefore not be confused with the so-called *non-informative prior pdf's* $f_I(\boldsymbol{\theta}_{1,2}|\boldsymbol{\theta}_{2,1})$ and $f_I(\boldsymbol{\theta})$ (see also Section 4.4, below). The properties of the consistency factors are extensively discussed in the next section.

4. THE CONSISTENCY FACTORS

4.1 General properties of the consistency factors

According to Proposition 3, for a consistent assignment of inverse probability distributions, the appropriate consistency factors $\zeta_{I,\Theta}(\theta)$ and $\zeta_{I,\Theta_{1,2}|\Theta_{2,1}}(\theta_1, \theta_2)$ need be uniquely determined. In what follows, we discuss some of the properties of the consistency factors that will be invoked during their determination.

PROPERTY 1 (Uniqueness). *A consistency factors $\zeta_{I,\Theta}(\theta)$ can only be determined up to a factor $\chi_{I,\Theta}(\mathbf{x}_{1,2})$ that is an arbitrary function of $\mathbf{x}_{1,2}$. Also, $\zeta_{I,\Theta_{1,2}|\Theta_{2,1}}(\theta_1, \theta_2)$ is determined only up to an arbitrary multiplier $\chi_{I,\Theta_{1,2}|\Theta_{2,1}}(\mathbf{x}_{1,2}, \theta_{2,1})$.*

Proof. Multiplying $\zeta_{I,\Theta}(\theta)$ by $\chi_{I,\Theta}(\mathbf{x}_{1,2})$ results in multiplying $\eta_{I,\Theta}(\mathbf{x}_{1,2})$ by the same factor, such that the factor cancels in the ratio $\zeta_{I,\Theta}(\theta)/\eta_{I,\Theta}(\mathbf{x}_{1,2})$. Identical arguments apply when $\zeta_{I,\Theta_{1,2}|\Theta_{2,1}}(\theta_1, \theta_2)$ is multiplied by $\chi_{I,\Theta_{1,2}|\Theta_{2,1}}(\mathbf{x}_{1,2}, \theta_{2,1})$. \square

PROPERTY 2 (Sign). *A consistency factor $\zeta_{I,\Theta}(\theta)$ is either positive or negative on the parameter space V_Θ , and so is $\zeta_{I,\Theta_{1,2}|\Theta_{2,1}}(\theta_1, \theta_2)$ on $V_{\Theta_{1,2}|\Theta_{2,1}}$.*

Proof. The normalization factors $\eta_{I,\Theta}(\mathbf{x}_{1,2})$ are either positive or negative, and the pdf's $f_I(\theta|\mathbf{x}_{1,2})$ and $f_I(\mathbf{x}_{1,2}|\theta)$ are non-negative, such that $\zeta_{I,\Theta}(\theta)$ must be of the same sign as $\eta_{I,\Theta}(\mathbf{x}_{1,2})$, i.e., either positive or negative for all $\theta \in V_\Theta$. The same holds true for $\eta_{I,\Theta_{1,2}|\Theta_{2,1}}(\mathbf{x}_{1,2}, \theta_{2,1})$, $f_I(\theta_{1,2}|\theta_{2,1}, \mathbf{x}_{1,2})$, $f_I(\mathbf{x}_{1,2}|\theta_1, \theta_2)$ and $\zeta_{I,\Theta_{1,2}|\Theta_{2,1}}(\theta_1, \theta_2)$. \square

PROPERTY 3 (Transformations). *Suppose that the premises of Proposition 3 are fulfilled such that pdf's $f_I(\theta|\mathbf{x}_{1,2})$ and $f_I(\mathbf{x}_{1,2}|\theta)$ are related according to (44). Let, in addition, $(\bar{\mathbf{s}}, \mathbf{s}) : (\Theta, \mathbf{X}) \rightarrow (\bar{\mathbf{s}} \circ \Theta, \mathbf{s} \circ \mathbf{X}) \equiv (\Lambda, \mathbf{Y})$ be a differentiable transformation with non-vanishing Jacobians $|\partial_\lambda \bar{\mathbf{s}}(\lambda)|$ and $|\partial_{\mathbf{x}_{1,2}} \mathbf{s}(\mathbf{x}_{1,2})|$ for all θ and $\mathbf{x}_{1,2}$ for which $f_I(\mathbf{x}_{1,2}|\theta)$ is positive. Then, the consistency and the normalization factors that relate $f_{I'}(\lambda|\mathbf{y}_{1,2})$ and $f_{I'}(\mathbf{y}_{1,2}|\lambda)$ read*

$$(45) \quad \zeta_{I',\Lambda}(\lambda) = \chi_{I',\Lambda} \zeta_{I,\Theta}[\bar{\mathbf{s}}^{-1}(\lambda)] |\partial_\lambda \bar{\mathbf{s}}^{-1}(\lambda)|$$

and

$$(46) \quad \eta_{I',\Lambda}(\mathbf{y}_{1,2}) = \chi_{I',\Lambda} \eta_{I,\Theta}[\mathbf{s}^{-1}(\mathbf{y}_{1,2})] |\partial_{\mathbf{y}_{1,2}} \mathbf{s}^{-1}(\mathbf{y}_{1,2})|.$$

Similarly, for $f_I(\theta_{1,2}|\theta_{2,1}, \mathbf{x}_{1,2})$ and $f_I(\mathbf{x}_{1,2}|\theta_1, \theta_2)$,

$$(47) \quad \zeta_{I',\Lambda_{1,2}|\Lambda_{2,1}}(\lambda_1, \lambda_2) = \chi_{I',\Lambda_{1,2}|\Lambda_{2,1}} \times \zeta_{I,\Theta_{1,2}|\Theta_{2,1}}[\bar{\mathbf{s}}_1^{-1}(\lambda_1), \bar{\mathbf{s}}_2^{-1}(\lambda_2)] |\partial_{\lambda_{1,2}} \bar{\mathbf{s}}_{1,2}^{-1}(\lambda_{1,2})|$$

and

$$(48) \quad \eta_{I',\Lambda_{1,2}|\Lambda_{2,1}}(\mathbf{y}_{1,2}, \lambda_{2,1}) = \chi_{I',\Lambda_{1,2}|\Lambda_{2,1}} \times \eta_{I,\Theta_{1,2}|\Theta_{2,1}}[\mathbf{s}^{-1}(\mathbf{y}_{1,2}), \bar{\mathbf{s}}_{2,1}^{-1}(\lambda_{2,1})] |\partial_{\mathbf{y}_{1,2}} \mathbf{s}^{-1}(\mathbf{y}_{1,2})|$$

are the transformations of the consistency and the normalization factors that are induced by the transformations $(\bar{\mathbf{s}}_1, \bar{\mathbf{s}}_2, \mathbf{s}) : (\Theta_1, \Theta_2, \mathbf{X}) \rightarrow (\bar{\mathbf{s}}_1 \circ \Theta_1, \bar{\mathbf{s}}_2 \circ \Theta_2, \mathbf{s} \circ \mathbf{X}) \equiv (\Lambda_1, \Lambda_2, \mathbf{Y})$ of the random variable $(\Theta_1, \Theta_2, \mathbf{X})$.

Proof. Combining equations (18) and (39) results in

$$f_{I'}(\lambda_{1,2}|\lambda_{2,1}, \mathbf{y}_{1,2}) = \frac{\zeta_{I,\Theta_{1,2}|\Theta_{2,1}}[\bar{\mathbf{s}}_1^{-1}(\lambda_1), \bar{\mathbf{s}}_2^{-1}(\lambda_2)] |\partial_{\lambda_{1,2}} \bar{\mathbf{s}}_{1,2}^{-1}(\lambda_{1,2})|}{\eta_{I,\Theta_{1,2}|\Theta_{2,1}}[\mathbf{s}^{-1}(\mathbf{y}_{1,2}), \bar{\mathbf{s}}_{2,1}^{-1}(\lambda_{2,1})] |\partial_{\mathbf{y}_{1,2}} \mathbf{s}^{-1}(\mathbf{y}_{1,2})|} \times f_{I'}(\mathbf{y}_{1,2}|\lambda_1, \lambda_2),$$

which, when compared to the relation

$$f_{I'}(\lambda_{1,2}|\lambda_{2,1}, \mathbf{y}_{1,2}) = \frac{\zeta_{I',\Lambda_{1,2}|\Lambda_{2,1}}(\lambda_1, \lambda_2)}{\eta_{I',\Lambda_{1,2}|\Lambda_{2,1}}(\mathbf{y}_{1,2}, \lambda_{2,1})} f_{I'}(\mathbf{y}_{1,2}|\lambda_1, \lambda_2),$$

implied by Proposition 3, yields (47) and (48). In the same way, (45) and (46) are obtained if (39) is replaced by (38). \square

For invariant families I of direct probability distributions, equations (45) and (47) reduce to functional equations

$$(49) \quad \zeta_{I,\Theta}(\theta) = \chi_{I,\Theta}(a) \zeta_{I,\Theta}[\bar{\mathbf{g}}_a^{-1}(\theta)] |\partial_\theta \bar{\mathbf{g}}_a^{-1}(\theta)|$$

and

$$(50) \quad \zeta_{I,\Theta_{1,2}|\Theta_{2,1}}(\theta_1, \theta_2) = \chi_{I,\Theta_{1,2}|\Theta_{2,1}}(a) \times \zeta_{I,\Theta_{1,2}|\Theta_{2,1}}[\bar{\mathbf{g}}_{a,1}^{-1}(\theta_1), \bar{\mathbf{g}}_{a,2}^{-1}(\theta_2)] |\partial_{\theta_{1,2}} \bar{\mathbf{g}}_{a,1,2}^{-1}(\theta_{1,2})|$$

for the consistency factors $\zeta_{I,\Theta}(\theta)$ and $\zeta_{I,\Theta_{1,2}|\Theta_{2,1}}(\theta_1, \theta_2)$, respectively. It should be noticed that the usual multipliers $\chi_{I,\Theta}$ and $\chi_{I,\Theta_{1,2}|\Theta_{2,1}}$, up to which the two consistency factors are uniquely determined (Property 1), may depend on the parameters a of the transformations (on the group elements a), i.e., the consistency factors for the parameters of invariant parametric families of direct probability distributions are to be *relatively invariant* under $\bar{\mathcal{G}}$.

Apart from the invariance of the consistency factors, invariance of a family I of direct distributions under a group \mathcal{G} also implies invariance of the family of the corresponding inverse distributions under the induced group $\bar{\mathcal{G}}$. Let, for example, I be an invariant parametric family of continuous direct probability distributions of a scalar random variable X , whose scalar parameter is denoted by Θ . Then, according to (7),

$$(51)$$

$$F_I(\theta|x) = \begin{cases} F_I(\bar{h}(a^{-1}, \theta)|h(a^{-1}, x)) & ; \bar{g}'_a(\theta) > 0 \\ 1 - F_I(\bar{h}(a^{-1}, \theta)|h(a^{-1}, x)) & ; \bar{g}'_a(\theta) < 0 \end{cases}.$$

4.2 Invariance under discrete groups of transformations

Under what circumstances functional equations (49) and (50) lead to unique solutions $\zeta_{I,\Theta}(\theta)$ and $\zeta_{I,\Theta_{1,2}|\Theta_{2,1}}(\theta_1, \theta_2)$?

EXAMPLE 7 (Parity). Let a parametric family of continuous direct probability distributions be invariant under a discrete group \mathcal{G} of transformations $g_a : X \rightarrow aX$ with $\bar{g}_a : \Theta \rightarrow a\Theta$ being the corresponding transformations from the induced group, where the underlying group G consists of two elements, $a = \pm 1$. That is, the distributions from the considered family have (positive) *parity* under simultaneous inversions of the spaces of X and Θ . By combining $\zeta_{I,\Theta}(\theta) = \chi_{I,\Theta}(a) \zeta_{I,\Theta}[\bar{g}_a^{-1}(\theta)]$ and $\zeta_{I,\Theta}[\bar{g}_a^{-1}(\theta)] = \chi_{I,\Theta}(a) \zeta_{I,\Theta}\{g_a^{-1}[\bar{g}_a^{-1}(\theta)]\}$ and setting $a = -1$ we obtain $\zeta_{I,\Theta}(\theta) = \chi_{I,\Theta}(-1) \zeta_{I,\Theta}(-\theta)$ and $\zeta_{I,\Theta}(\theta) = [\chi_{I,\Theta}(-1)]^2 \zeta_{I,\Theta}(\theta)$, such that $[\chi_{I,\Theta}(-1)]^2 = 1$. When inability of $\zeta_{I,\Theta}(\theta)$ to switch sign is invoked (Property 2), this further implies $\zeta_{I,\Theta}(-\theta) = \zeta_{I,\Theta}(\theta)$. That is, $\zeta_{I,\Theta}(\theta)$ must have positive parity under the inversion $\Theta \rightarrow -\Theta$, but apart from this, it can take any form and so in this case equation (49) does not lead to unique solution.

It is not difficult to understand that this is a common feature of all solutions based on invariance of parametric families under *discrete* groups. If the symmetry group is discrete, the spaces of X and Θ break up in intervals, the so-called *fundamental regions* or *domains* of the group (Wigner (1959), § 19.1, p. 210; Jaynes (2003), § 10.9, p. 332), with no connections in terms of group transformations within the points of the same interval. We are then free to choose the form of $\zeta_{I,\Theta}(\theta)$ in one of these intervals (e.g., we can choose $\zeta_{I,\Theta}(\theta)$ for the positive values of θ in the above example), hence the invariance of a family I under a discrete group \mathcal{G} alone does not lead to a unique form of the corresponding consistency factor. The argument applies, for example, for all parametric families of discrete probability distributions.

4.3 Consistency factors and invariance under Lie groups

Let $\mathcal{G} = \{g_a : \mathbb{R} \rightarrow \mathbb{R}; a \in G\}$ be a group and G be a one-dimensional Lie group. Then, according to Proposition 2, on the subspace $\tilde{V}_X \subseteq V_X$ with non-vanishing derivative (27), every \mathcal{G} -invariant parametric family I of continuous direct probability distributions is necessarily isomorphic to a location-scale family I' with the realization $\sigma = 1$ of the scale parameter Θ_2 . Since the fundamental domain of the group $\bar{\mathcal{G}}$ of translations on the real axis consists of a single point, the space of all possible realizations of a location parameter is a *homogenous space for the group* (i.e., the space is said to be a *single $\bar{\mathcal{G}}$ -orbit*).

The implications of Proposition 2 may be extended to the subspaces $V_X - \tilde{V}_X$:

PROPOSITION 4. Let $\mathcal{G} = \{g_a : a \in G\}$ be a group of transformations $g_a : \mathbb{R} \rightarrow \mathbb{R}$ and G be a one-dimensional Lie group. Suppose, in addition, that a parametric family I of continuous direct probability distributions for a scalar random variable X is \mathcal{G} -invariant, that the action of G on \mathbb{R} is not identically trivial on entire V_X , and that the corresponding cdf's $F_I(x|\lambda)$ are differentiable in λ . Then, for a realization $x \in V_X - \tilde{V}_X \subset V_X$ with vanishing derivative (27), the inverse probability distribution whose cdf $F_I(\lambda|x)$ is differentiable in x , cannot be assigned. (Existence of derivatives $\partial_x F_I(x|\lambda)$ and $\partial_\lambda F_I(\lambda|x)$ is assured by Definition 5.)

EXAMPLE 8. Let $I_\mu \equiv \{P_{I,(\mu,\sigma)} : (\mu,\sigma) \in (\mu, \mathbb{R}^+)\}$ be a sub-family of a continuous location-scale family I that corresponds to the value of the location parameter Θ_1 being fixed to μ . By transformation $X \longrightarrow X - \mu \equiv Y$, every cdf $F_{I_\mu}(x|\mu, \sigma)$ from I_μ is reduced to

$$F_{I'_\mu}(y|\mu, \sigma) = F_{I_\mu}(y + \mu|\mu, \sigma) = \Phi\left(\frac{y}{\sigma}\right),$$

where $y \equiv x - \mu$. The probability distribution for the random variable Y thus belongs to the family I'_μ that is invariant under transformations $g_a : Y \longrightarrow aY$ and $\bar{g}_a : \Theta_2 \longrightarrow a\Theta_2$ for all $a \in \mathbb{R}^+$. Since the derivative $\partial_a h(a^{-1}, y)|_{a=e} = y$ vanishes for $y = 0$, the inverse probability distribution for the scale parameter Θ_2 given $y = 0$ (or, equivalently, given $x = \mu$) does not exist.

In order to assign an inverse probability distribution to a scalar parameter of a family that is invariant under a group \mathcal{G} that is underlain by a one dimensional Lie group G it therefore suffices to determine the consistency factor $\zeta_{I, \Theta_1}(\mu) \equiv \zeta_{I, \Theta_1|\sigma=1}(\mu, \sigma)$, which can subsequently be transformed, by means of (45), to the corresponding consistency factor $\zeta_{I', \Lambda}(\lambda)$ for the original parameter Λ . A location-scale family $I_{\sigma=1} = \{Pr_{I',(\mu,\sigma)} : (\mu, \sigma) \in (\mathbb{R}, 1)\}$ of continuous direct probability distributions with the fixed value $\sigma = 1$ of the scale parameter is a subset of the location-scale family $I = \{Pr_{I,(\mu,\sigma)} : (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}^+\}$ that is invariant under the group \mathcal{G} (25). Given a location-scale family I , the functional equation (49) for the consistency factor $\zeta_{I, \Theta_1|\sigma}(\mu, \sigma)$ therefore reduces to

$$(52) \quad \zeta_{I, \Theta_1|\sigma}(\mu, \sigma) = h(a_1, a_2) \zeta_{I, \Theta_1|\sigma}[(\mu - a_1)/a_2, \sigma/a_2],$$

$\mu, a_1 \in \mathbb{R}$ and $\sigma, a_2 \in \mathbb{R}^+$, where $h(a_1, a_2) \equiv \chi_{I, \Theta_1|\sigma}(a_1, a_2)/a_2$.

LEMMA 4. *The solution $\zeta_{I, \Theta_1|\sigma}(\mu, \sigma)$ of equation (52) is a function of σ alone, say $\Omega(\sigma)$.*

Since $\zeta_{I, \Theta_1|\sigma}(\mu, \sigma)$ is uniquely determined only up to a factor $\chi_{I, \Theta_1|\sigma}(x_{1,2}, \sigma)$ (Property 1), $\Omega(\sigma)$ may be, without loss of generality, set to unity, such that

$$(53) \quad \zeta_{I, \Theta_1|\sigma}(\mu, \sigma) = 1,$$

regardless the explicit family I of direct probability distributions, as well as the realization σ of the scale parameter.

By using the same arguments as for $\zeta_{I, \Theta_1|\sigma}(\mu, \sigma)$ we find that a consistency factor $\zeta_{I, \Theta_2|\mu}(\mu, \sigma)$ is also a function of σ only, say $\zeta_{I, \Theta_2}(\sigma) \equiv \zeta_{I, \Theta_2|\mu}(\mu, \sigma)$. The inverse probability distribution for the scale parameter Θ_2 , given $\Theta_1 = \mu$ and $X_1 = x_1 = \mu$, does not exist (Example 8), while for $x_1 \gtrless \mu$ the pdf $f_I(\sigma|\mu, x_1)$ can be expressed in terms of $f_{I^\pm}(x_1|\mu, \sigma)$ (Section 2.2):

$$\begin{aligned} f_I(\sigma|\mu, x_1) &= \frac{\zeta_{I, \Theta_2|\mu}(\mu, \sigma) f_I(x_1|\mu, \sigma)}{\eta_{I, \Theta_2|\mu}(x_1, \mu)} \\ &= \frac{\zeta_{I, \Theta_2}(\sigma) f_{I^\pm}(x_1|\mu, \sigma)}{\eta_{I^\pm, \Theta_2|\mu}(x_1, \mu)}, \end{aligned}$$

where $\eta_{I^\pm, \Theta_2|\mu}(x_1, \mu) \equiv \eta_{I, \Theta_2|\mu}(x_1, \mu)/c_\pm$. By equation (22), every pdf $f_{I^\pm}(x_1|\mu, \sigma)$ is reducible to $f_{I^\pm}(y_1|\lambda_1, \lambda_2 = 1)$, such that

$$f_{I'}(\lambda_1|\lambda_2 = 1, y_1) = \frac{\zeta_{I', \Lambda_1}(\lambda_1) f_{I^\pm}(y_1|\lambda_1, \lambda_2 = 1)}{\eta_{I^\pm, \Lambda_1|\lambda_2=1}(y_1, \lambda_2 = 1)}$$

holds true and $\zeta_{I', \Lambda_1}(\lambda_1) \equiv \zeta_{I', \Lambda_1|\lambda_2=1}(\lambda_1, \lambda_2) = 1$, where $y_1 \equiv \ln\{\pm(x_1 - \mu)\}$ and $\lambda_1 \equiv \ln \sigma \equiv \bar{s}(\sigma)$. Since, according to equation (45),

$$\zeta_{I', \Lambda_1}(\lambda_1) = \zeta_{I, \Theta_2}[\bar{s}^{-1}(\lambda_1)] |[\bar{s}^{-1}(\lambda_1)]'|$$

must also hold,

$$(54) \quad \zeta_{I, \Theta_2|\mu}(\mu, \sigma) = \sigma^{-1}$$

is the general form of the consistency factor $\zeta_{I, \Theta_2|\mu}(\mu, \sigma)$, again regardless the explicit location-scale family I of direct probability distributions and the realization μ of the location parameter.

According to Proposition 3, an inverse pdf $f_I(\mu, \sigma|x_1, x_2)$ for the parameters Θ_1 and Θ_2 of a location-scale family I must be expressible as

$$f_I(\mu, \sigma|x_1, x_2) = \frac{\zeta_{I, \Theta}(\mu, \sigma) f_I(x_1, x_2|\mu, \sigma)}{\eta_{I, \Theta}(x_1, x_2)}.$$

For the same reasons as $\zeta_{I, \Theta_1|\sigma}(\mu, \sigma)$ (Lemma 4), $\zeta_{I, \Theta}(\mu, \sigma)$ must also be a function of σ alone, say $\Xi(\sigma)$, while the product rule (40) implies factorizability of $f_I(\mu, \sigma|x_1, x_2)$,

$$(55) \quad f_I(\mu, \sigma|x_1, x_2) = f_I(\sigma|\mu, x_1, x_2) f_I(\mu|x_1, x_2),$$

where, according to Bayes' Theorem (41),

$$\begin{aligned} f_I(\sigma|\mu, x_1, x_2) &= \frac{f_I(\sigma|\mu, x_1) f_I(x_2|\mu, \sigma)}{f_I(x_2|\mu, x_1)} \\ &= \frac{\zeta_{I, \Theta_2|\mu}(\mu, \sigma) f_I(x_1, x_2|\mu, \sigma)}{\eta_{I, \Theta_2|\mu}(x_1, \mu) f_I(x_2|\mu, x_1)}. \end{aligned}$$

Hence,

$$\frac{\Xi(\sigma)}{\eta_{I, \Theta}(x_1, x_2)} = \frac{\sigma^{-1} f_I(\mu|x_1, x_2)}{\eta_{I, \Theta_2|\mu}(x_1, \mu) f_I(x_2|\mu, x_1)}.$$

must hold, finally implying

$$(56) \quad \Xi(\sigma) = \zeta_{I, \Theta}(\mu, \sigma) = \sigma^{-1}.$$

The findings of the present subsection can thus be recapitulated as follows:

PROPOSITION 5. *The consistency factors $\zeta_{I, \Theta_1|\sigma}(\mu, \sigma)$, $\zeta_{I, \Theta_2|\mu}(\mu, \sigma)$ and $\zeta_{I, \Theta}(\mu, \sigma)$ for the parameters of location-scale families of continuous direct probability distributions read $\zeta_{I, \Theta_1|\sigma}(\mu, \sigma) = 1$ and $\zeta_{I, \Theta_2|\mu}(\mu, \sigma) = \zeta_{I, \Theta}(\mu, \sigma) = \sigma^{-1}$.*

4.4 On integrability and on uniqueness of the consistency factors

It is easily verified that normalizability of pdf's (21) from location-scale families guarantees also normalizability (integrability) of all the pdf's that were involved in the foregoing derivations of the consistency factors. No requirement concerning integrability, however, has ever been imposed to consistency factors themselves. Moreover, it is evident that consistency factors $\zeta_{I, \Theta_1|\sigma}(\mu, \sigma)$ (53), defined on the entire real axis, are not integrable, implying that none of the consistency factors for scalar parameters of parametric families that are invariant under the action of a one-dimensional Lie group, is integrable.

Let $\zeta_{I, \Theta}(\theta)$ be a non-integrable consistency factor for a parameter θ from a family I of continuous direct probability distributions. Suppose for a moment that apart from the conditional pdf's $f_I(\mathbf{x}|\theta)$ and $f_I(\theta|\mathbf{x})$, there also exist the non-informative prior pdf $f_I(\theta)$ and the joint pdf $f_I(\theta, \mathbf{x})$. Then, there exists an unconditional predictive pdf $f_I(\mathbf{x})$ (see, for example, Shao (1999), § 4.1.1, Theorem 4.1, p. 194), such that

$$(57) \quad f_I(\theta|\mathbf{x}) = \frac{f_I(\theta) f_I(\mathbf{x}|\theta)}{f_I(\mathbf{x})}.$$

But apart from Bayes' Theorem (57), $f_I(\theta|\mathbf{x})$ is also subjected to Proposition 3, implying that $f_I(\theta)$ and $\zeta_{I, \Theta}(\theta)$ are equal up to an arbitrary multiplication constant. Since then $f_I(\theta)$ is not integrable, the non-informative pdf $f_I(\theta)$ does not exist, and consequently, neither do exist $f_I(\theta, \mathbf{x})$ and the underlying probability space (Ω, Σ, P) . The pdf's $f_I(\mathbf{x}|\theta)$ and $f_I(\theta|\mathbf{x})$ therefore represent an extension of the concept of the conditional probability distribution that was introduced in Subsection 2.1.

Since every consistency factor is determined only up to an arbitrary multiplicative factor (Property 1), infinitely many different consistency factors for a parameter from a particular parametric family exist. Nevertheless, unlike non-unique non-informative prior probability distributions (recall the assertions quoted in the introductory remarks), for a scalar parameter of a family of direct probability distributions whose invariance is associated to a one-dimensional Lie group, for example, the consistency factors are unique in that they all lead to the same inverse probability distribution.

4.5 Discussion

Above, the consistency factors were deduced exclusively by presuming existence of the inverse probability distributions and by making use of the invariance of the families of direct probability distributions that is related to Lie groups. The resulting set of the families with possible probabilistic parametric inference is limited: for example, for scalar random variables X and scalar parameters θ the probabilistic parametric inference is in this way restricted to location parameters (or to parameters that are reducible to location parameters by one-to-one transformations). On the other hand, several principles were proposed for determination of the non-informative prior probability distributions. Here, applicability of these principles for determination of the consistency factors is investigated in order to extend the domain of the probabilistic parametric inference.

For example, if adapted for determination of consistency factors, *Bayes' Postulate* (Bayes, 1763), also referred to as the *Laplace Principle of Insufficient Reason* (Laplace (1886), p. XVII), suggests that all consistency factors should be uniform. Clearly, this is inadmissible since in general the constant consis-

tency factors contradict expressions (45) and (47) for transformations of the consistency factors under reparameterizations.

A sophisticated version of the Principle of Insufficient Reason is referred to as the *Principle of Maximum Entropy*. In our context, the *information entropy* (Shannon (1948), § 6) reads

$$S \equiv - \int_{V_\Theta} \zeta_{I,\Theta}(\theta) \ln \{ \zeta_{I,\Theta}(\theta) \} d\theta ,$$

while the Principle of Maximum Entropy states (Jaynes (2003), § 11.3, pp. 350) that the consistency factor which maximizes the entropy represents the most honest description of what we know about the value of the inferred parameter. For compact parameter spaces V_Θ for which the above integral exists, the principle again results in constant consistency factors $\zeta_{I,\Theta}(\theta) = e^{-1}$. The factors are then flawed in the same way as the factors implied by Bayes' Postulate. Jaynes (2003, § 12.3, pp. 374-377) argues that the above expression for the entropy is inappropriate since it is not invariant under reparameterization and proposes a *Kullback-Leibler divergence* (also called *relative entropy*) to replace it:

$$S \equiv - \int_{V_\Theta} \zeta_{I,\Theta}(\theta) \ln \left\{ \frac{\zeta_{I,\Theta}(\theta)}{m(\theta)} \right\} d\theta ,$$

where $m(\theta)$ is the *reference measure function*. Due to the unknown form of the latter, however, maximization of the relative entropy does not lead to unique consistency factors.

If *Jeffreys' general rule* is applied (Jeffreys (1946)), the consistency factors are determined *via* the determinant of the Fisher information matrix $\mathcal{I}_{I,\Theta}(\theta)$, $\zeta_{I,\Theta}(\theta) \propto \sqrt{\det [\mathcal{I}_{I,\Theta}(\theta)]}$, where the elements of the matrix are given by

$$[\mathcal{I}_{I,\Theta}(\theta)]_{i,j} \equiv \int_{\mathbb{R}^n} \partial_{\theta_i} \ln \{ f_I(\mathbf{x}|\theta) \} \partial_{\theta_j} \ln \{ f_I(\mathbf{x}|\theta) \} f_I(\mathbf{x}|\theta) d^n \mathbf{x} .$$

The obtained consistency factors satisfy requirements (45) and (47) for transformations of the factors under reparameterization, but are flawed in another way. Let, for example, a probability distribution $N(\mu, \sigma)$ for a random variable X belong to the normal (or Gaussian) family (Stuart and Ord (2000), § 5.36, p. 191). Then, Jeffrey's general rule

yields the consistency factors $\zeta_{I,\Theta_1|\sigma}(\mu, \sigma) \propto 1$, $\zeta_{I,\Theta_2|\mu}(\mu, \sigma) \propto \sigma^{-1}$ and $\zeta_{I,\Theta}(\mu, \sigma) \propto \sigma^{-2}$, such that the resulting inverse probability distributions violate the product rule (55).

A modification of Jeffreys' general rule by Bernardo (1979) called the *reference prior approach* leads to violations of the same product rule (Bernardo (1979), § 3.3, pp. 118-119). Also, let X_1, \dots, X_n be independent random variables with identical probability distribution $N(\mu, \sigma)$. Since the normal family is a location-scale family of continuous distributions, the consistency factor (56) yields a unique posterior pdf $f_I(\bar{s}^{-1}(\lambda, \sigma)|\mathbf{x})$ for the parameter (Θ_1, Θ_2) of the distribution, whereas the posterior pdf for $(\Lambda, \Theta_2) \equiv \bar{s} \circ (\Theta_1, \Theta_2)$, $\Lambda \equiv \Theta_1/\Theta_2$, is obtained according to (45) (Property 3), $f_{I'}(\lambda, \sigma|\mathbf{x}) = f_I(\bar{s}^{-1}(\lambda, \sigma)|\mathbf{x}) |\partial_{(\lambda, \sigma)} \bar{s}^{-1}(\lambda, \sigma)| = \sigma f_I(\bar{s}^{-1}(\lambda, \sigma)|\mathbf{x})$. A unique $f_{I'}(\lambda, \sigma|\mathbf{x})$ further implies a unique marginal pdf $f_{I'}(\lambda|\mathbf{x})$,

$$\begin{aligned} f_{I'}(\lambda|\mathbf{x}) &= \int_0^\infty f_{I'}(\lambda, \sigma|\mathbf{x}) d\sigma \\ &\propto \exp \left\{ -n\lambda^2/2 \right\} \int_0^\infty u^n \exp \left\{ -\frac{u^2}{2} + r\lambda u \right\} du, \end{aligned}$$

$r \equiv (\sum x_i)/\sqrt{\sum x_i^2}$, while the reference prior approach leads to

$$\begin{aligned} f_{I'}(\lambda|\mathbf{x}) &\propto \\ &\frac{\exp \left\{ -n\lambda^2/2 \right\}}{\sqrt{1 + \lambda^2/2}} \int_0^\infty u^{n-1} \exp \left\{ -\frac{u^2}{2} + r\lambda u \right\} du \end{aligned}$$

(Bernardo (1979), § 5.1, pp. 122-123). In this way, since the two expressions for $f_{I'}(\lambda|\mathbf{x})$ are incompatible, inconsistency of the reference prior approach with the probabilistic parametric inference is once more demonstrated.

Invariance theory has played an important role in the theory of non-informative prior probability distributions (see, for example, Hartigan (1964); Jaynes (1968) and 2003, Chapter 12, pp. 372-396; Dawid et al. (1973), Section 2, pp. 195-199; Villegas (1977) and 1981; Eaton (1989); Kass and Wasserman (1996), § 3.2, pp. 1347-1348). Functional equations (49) and (50), for instance, correspond to what has been called the *Principle of Relative Invariance* (Hartigan, 1964). Since the relative invariance of the consistency factors is implied

immediately by the existence of the inverse probability distributions, the Principle of Relative Invariance, when applied to consistency factors, is redundant. Contrary to what is demonstrated above, it has also been believed that the Principle is insufficient to determine uniquely defined priors (consistency factors) (Hartigan (1964), § 4, p. 838 and § 10, p. 845; Villegas (1977), § 2, p. 454; Kass and Wasserman (1996), § 3.2, p. 1348).

If multipliers $\chi_{I,\Theta}(a)$ and $\chi_{I,\Theta_{1,2}|\Theta_{2,1}}(a)$ are set to unity, equations (49) and (50) lead to *inner* (or *form invariant*) *consistency factors* (Villegas (1977); Harney (2003), § 2.3, pp. 11-12 and § 6.3, pp. 53-54). Since, however, the form invariant consistency factors for location-scale families, $\zeta_{I,\Theta_1|\sigma}(\mu, \sigma) \propto 1$, $\zeta_{I,\Theta_2|\mu}(\mu, \sigma) \propto \sigma^{-1}$ and $\zeta_{I,\Theta}(\mu, \sigma) \propto \sigma^{-2}$, lead to a violation of the product rule (55), the *Principle of Form Invariance* is inconsistent with the probabilistic parametric inference.

When a parameter space V_Θ of a family I is identical to the symmetry group G of the family, every realization θ of the parameter Θ identifies both an element of the family I and an element in G . If, in addition, the left action $\bar{I} : G \times G \longrightarrow G$ coincides with the composition of the group elements a and θ , $\bar{I}(a, \theta) \equiv a \circ \theta$, the form invariant consistency factors $\zeta_{I,\Theta}(\theta)$ are called the *left Haar consistency factors*, where *left Haar* is due to the multiplication of θ by a from the left and due to the fact that $\nu_{l,H}(d^m \theta) = \zeta_{I,\Theta}(\theta) d^m \theta$ leads to the left-invariant Haar measure

$$\nu_{l,H}(B) = \int_B \nu_{l,H}(d^m \theta)$$

on \mathcal{B}^m (Haar, 1933), i.e., $\nu_{l,H}(a \circ B) = \nu_{l,H}(B)$ for all $a \in G$ and $B \in \mathcal{B}^m$, where $a \circ B \equiv \{a \circ b : b \in B\}$. Likewise, when $\bar{k}(a, \theta) \equiv \theta \circ a$, the consistency factors that solve the functional equation $\zeta_{I,\Theta}(\theta) = \zeta_{I,\Theta}[\bar{k}(a^{-1}, \theta)] |\partial_\theta \bar{k}(a^{-1}, \theta)|$ are called *right Haar consistency factors* on G . When G is a *topological group*, e.g., a Lie group, both the left and the right Haar measures (consistency factors) exist and are unique, each up to a positive multiplication constant (Nachbin, 1965), but the two measures (consistency factors) need not coincide. For the location-scale families, for example, $\bar{k}[a, (\mu, \sigma)] = (\mu + a_1 \sigma, a_2 \sigma)$ induces the right Haar consistency factor $\zeta_{I,\Theta}(\mu, \sigma) = \sigma^{-1}$ which, in contrast to the

corresponding left Haar factor, does not lead to the violation of the product rule (55).

Several additional desirable properties are established for the right Haar consistency factors (see Section 5.2 below for an example). Nevertheless, Eaton and Sudderth (1993, 1999, 2002) showed that unless the symmetry groups are further restricted to, for example, amenable groups, the probability distributions based on the predictive pdf's that are obtained by applications of the right Haar consistency factors (priors) are not generally consistent with the probability axioms. We cannot tell though, whether or not the right-invariant consistency factors based on the restricted groups extend the collection of families for which the probabilistic parametric inference is possible.

In summary, except possibly for the principle that identifies consistency factors with the right Haar factors for the underlying symmetry group G , all the principles discussed are either redundant, inconsistent with the probabilistic parametric inference, or do not lead to unique consistency factors.

5. INTERPRETATIONS OF PROBABILITY DISTRIBUTIONS

Every axiomatic (abstract) theory admits, as is well known, of an unlimited number of concrete interpretations besides those from which it was derived. Thus we find applications in fields of science which have no relations to the concepts of random event and of probability in the precise meaning of these words.

Kolmogorov (1933), Chapter 1, p. 1.

5.1 Probability distributions, relative frequencies and degrees of belief

So far, a mathematical theory of probabilistic parametric inference has been discussed. In the present section, however, two concepts of probability distributions are introduced that link the mathematical theory to an external world of measurable phenomena: the *concept of relative frequencies* in repeated trials, and the *concept of degrees of belief* in *hypotheses* or *propositions* (i.e., in statements that can be either true or false) concerning values of inferred parameters of parametric families.

Suppose an experiment is repeated under identical conditions, but the outcomes vary from one repetition of the experiment to another. If a numerical

characteristic assigned to the outcomes of the experiment follows no describable deterministic pattern, the experiment is called *random experiment*, the outcomes of the experiment are called *random events*, while the underlying process of such an experiment is called *random process*. Let random events be mutually independent. Then, within the frequency interpretation of probability distributions, the direct probability distribution for a random variable \mathbf{X} , linked to the experiment, is assumed to coincide with the long term distribution of relative frequencies of particular outcomes of the experiment,

$$F_I(\mathbf{x}|\boldsymbol{\theta}) = \lim_{N \rightarrow \infty} \frac{N_{\mathbf{X} \leq \mathbf{x}}}{N},$$

where N is the total number of repetitions of the experiment and $N_{\mathbf{X} \leq \mathbf{x}}$ is the number of the repetitions with outcomes whose numerical characteristic is less-or-equal to \mathbf{x} . Henceforth, the frequency interpretation of direct probability distributions is assumed.

Inverse probability distributions, on the other hand, are used to express one's degrees of belief that, given a (finite) recorded sequence $\mathbf{x}_1, \mathbf{x}_2, \dots$ of realizations of independent random variables $\mathbf{X}_1, \mathbf{X}_2, \dots$ with an identical probability distribution from a parametric family I , the so-called *true value* of the parameter $\boldsymbol{\Theta}$ of the family (i.e., the value of the parameter that uniquely determines the true limiting frequency distribution of the realizations) lies within a certain region of the parameter space. Several strong arguments exist for inverse probability distributions being the ideal for parametric inferences, like, for example, the so-called Dutch Book Theorem, emerging from the work of Ramsey (1931, Chapter VII, pp. 156-198), de Finetti (1931, 1937), Shimony (1955) and Kemeny (1955), and Cox's Theorem (Cox, 1946). For a concise review of the two Theorems see, for example, Paris (1994), Chapter 3, pp. 19-33.

While being identical objects from a mathematical perspective, the direct and the inverse probability distributions obviously have different interpretations. Contrary to the distribution of realizations of random variables \mathbf{X}_i , in most situation the realization of a parameter – the inferred true value of $\boldsymbol{\Theta}$ – is unknown but fixed. Several authors overlooked this important difference

between the frequency distributions and the distributions of someone's beliefs (see, for example, Lehmann (1986), § 1.6, p. 14; Shao (1999), § 7.1.3, p. 431; Casella and Berger (2002), § 7.2.3, p. 324 and § 9.2.4, pp. 435-436; Harney (2003), § 2.5, p. 18). It should be noticed, however, that the developed theory of probabilistic parametric inference still provides verifiable predictions in terms of relative frequencies of confidence intervals, covering the true value of the parameter (see Section 5.2, below). The theory is then both operational and objective.

5.2 Calibration

DEFINITION 16 (Confidence intervals). *Let $f_I(\theta|x)$ be a pdf of a probability distribution for a scalar parameter Θ , $V_\Theta = (\theta_a, \theta_b)$, given realization x of a scalar random variable X from a parametric family I . A confidence interval $(\theta_1(x), \theta_2(x)) \subseteq V_\Theta$ is defined via the system of equations*

$$Pr_I(\theta_a, \theta_1|x) = \int_{\theta_a}^{\theta_1} f_I(\theta|x) d\theta = \alpha$$

and

$$Pr_I((\theta_1, \theta_2)|x) = \int_{\theta_1}^{\theta_2} f_I(\theta|x) d\theta = \delta,$$

where $\delta \in [0, 1]$ and $\alpha \in [0, 1 - \delta]$. The number δ is called the probability content of the interval.

Higher dimensional *confidence regions*, e.g., *m*-dimensional *confidence rectangles* ($m \geq 2$), for vector-parameters are defined in a similar way.

DEFINITION 17 (Calibration). *Let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be a set of realizations of independent continuous random variables $\mathbf{X}_1, \dots, \mathbf{X}_n$ from a parametric family I of direct probability distributions. The inverse probability distributions, assigned to the inferred parameter $\boldsymbol{\Theta}$ of the family I , given realizations \mathbf{x}_i , are called calibrated if, in the limit $n \rightarrow \infty$, the coverage of the corresponding confidence regions (i.e., the relative frequency of the regions that cover the true values of the inferred parameter) coincides with the probability content δ of the region.*

Calibration of probability distributions for inferences about location and scale parameters is guaranteed by the fact that the consistency factors $\zeta_{I, \theta_1|\sigma}(\mu, \sigma)$, $\zeta_{I, \theta_2|\mu}(\mu, \sigma)$ and $\zeta_{I, \theta}(\mu, \sigma)$, determined in Subsection 4.3, coincide with the right

Haar factors for the group \mathbb{R} for summations, for the group \mathbb{R}^+ for multiplications, and for the group $\mathbb{R} \times \mathbb{R}^+$ for operations (24), respectively Stein (1965); Chang and Villegas (1986). That is to say, the resulting confidence regions coincide with the so-called *classical confidence regions*, first propounded by Neyman (1937). It should be noticed that this holds true even if the true value of the inferred parameter arbitrarily varies from realization of one random variable to another.

It can further be shown that the consistency factors for location and scale parameters, determined in Subsection 4.3, provide for a simple frequency interpretation of the predictive distributions.

To relate probabilistic parametric inference to another concept – that of the *fiducial inference* – let $F_I(x|\lambda)$ be a cdf for a continuous one-dimensional random variable X that is either strictly increasing or strictly decreasing in a scalar parameter λ . Then, a sufficient condition for an inverse probability distribution to be calibrated – the so-called *fiducial condition* by Fisher (1956, § 3.6, p. 70) – reads:

$$(58) \quad f_I(\lambda|x) = |\partial_\lambda F_I(x|\lambda)|.$$

Observe that for the inverse pdf's, assigned to location and scale parameters by using the consistency factors (53) and (54), the condition (58) is satisfied. Also, it is easily shown that congruence with the fiducial condition is preserved under updating that is made in accordance with Bayes' Theorem.

Conformity with the fiducial condition (58) is invariant under one-to-one transformations $Y \equiv s \circ X$ and $\Theta \equiv \bar{s} \circ \Lambda$ with non-vanishing derivatives $\bar{s}'(\theta)$:

$$\begin{aligned} f_{I'}(\theta|y) &= f_I(\bar{s}^{-1}(\theta)|s^{-1}(y)) |[\bar{s}^{-1}(\theta)]'| \\ &= |\partial_{\bar{s}^{-1}(\theta)} F_I(\bar{s}^{-1}(\theta)|s^{-1}(y)) [\bar{s}^{-1}(\theta)]'| \end{aligned}$$

and therefore

$$f_{I'}(\theta|y) = |\partial_\theta F_{I'}(\theta|y)|,$$

where the last equality is due to equation

$$F_{I'}(\theta|y) = \begin{cases} F_I(\bar{s}^{-1}(\theta)|s^{-1}(y)) & ; \bar{s}'(\theta) > 0 \\ 1 - F_I(\bar{s}^{-1}(\theta)|s^{-1}(y)) & ; \bar{s}'(\theta) < 0 \end{cases}$$

that follows immediately from the definition of the inverse cdf's and from equation (7). In addition, by

combining equation (44) from Proposition 3 with the above fiducial condition we obtain:

$$(59) \quad \zeta_{I,\Lambda}(\lambda) \partial_x F_I(x|\lambda) \pm \eta_{I,\Lambda}(x) \partial_\lambda F_I(x|\lambda) = 0,$$

where the upper (lower) sign stands for cdf's which are strictly decreasing (increasing) in λ . By defining $H(x, \lambda) \equiv s(x) \mp \bar{s}(\lambda)$, with $s(x)$ and $\bar{s}(\lambda)$ being related to $\zeta_{I,\Lambda}(\lambda)$ and $\eta_{I,\Lambda}(x)$ as $s'(x) \equiv \eta_{I,\Lambda}(x)$ and $\bar{s}'(\lambda) \equiv \zeta_{I,\Lambda}(\lambda)$, functional equation (59) can be reduced to (30). Recall that the most general solution $F_I(x|\lambda)$ of equation (30) implies existence of a cdf $F_{I'}(y|\mu)$ for $Y \equiv s \circ X$ from a location-scale family I' with $\mu \equiv \pm \bar{s}(\lambda)$ being a realization of the location parameter $\Theta_1 \equiv \bar{s} \circ \Lambda$, whereas the scale parameter Θ_2 of the family I' is set to 1. That is, the fiducial condition (58) and the requirement (44) of Proposition 3 combined imply reducibility of an inferred parameter to a location parameter. (Lindley (1958) obtained the same result by combining the calibration condition (58) and Bayes' Theorem (57).) For scalar parameters, the consistency factors that were deduced on the basis of invariance of parametric families under the action of one-dimensional Lie groups are therefore the only consistency factors for which the resulting inverse probability distributions satisfy the fiducial condition (58).

6. CONCLUSIONS

For scalar parameters, invariance of a parametric family of direct probability distributions under the action of a one-dimensional Lie group leads to unique inverse probability distributions. The concept of invariance is equivalent to the concept of fiducial distributions, combined with implications of Proposition 3: both concepts lead to identical inverse distributions and are applicable under the same conditions. When this is observed, the original idea of Bayes (1763) and Laplace (1886) of embedding parametric inference in the framework of probability theory becomes perfectly compatible with the concept of the classical confidence intervals (Neyman, 1937) and with the concept of the fiducial distributions (Fisher, 1935). Therefore, provided that adherents of the Bayesian schools of parametric inference are willing to give up the notion of non-informative prior probability distributions, while at the same time adherents of the frequentist schools are willing to adopt a broader concept of random variable that leads

to existence of inverse probability distributions, a reconciliations between different paradigms can be reached, probably the same kind of reconciliation that Kendall Kendall (1949) had in mind when he wrote: “Neither party can avoid ideas of the other in order to set up and justify a comprehensive theory.”

APPENDIX A: PROOFS OF PROPOSITIONS AND LEMMATA

A.1 Proof of Proposition 1

The left-hand side of (10) can be rewritten as

$$\begin{aligned}
 (60) \quad \tilde{\nu}_{\mathbf{1}_{\mathbf{Y}^{-1}(U)}}(S) &= \int_{\mathbf{Z}^{-1}(S)} \mathbf{1}_{\mathbf{Y}^{-1}(U)}(\omega) dP(\omega) \\
 &= \int_{\Omega} \mathbf{1}_{\mathbf{Y}^{-1}(U)}(\omega) \mathbf{1}_{\mathbf{Z}^{-1}(S)}(\omega) dP(\omega) \\
 &= \int_{\mathbb{R}^n \times \mathbb{R}^m} \mathbf{1}_U(\mathbf{y}) \mathbf{1}_S(\mathbf{z}) dPr_{\mathbf{X}}(\mathbf{y}, \mathbf{z}) \\
 &= \int_{U \times S} f_{\mathbf{X}}(\mathbf{y}, \mathbf{z}) d^n \mathbf{y} d^m \mathbf{z} \\
 &= \int_{U \times S} f_{\mathbf{X}}(\mathbf{z}) \frac{f_{\mathbf{X}}(\mathbf{y}, \mathbf{z})}{f_{\mathbf{X}}(\mathbf{z})} d^n \mathbf{y} d^m \mathbf{z} \\
 &= \int_S h(\mathbf{z}) f_{\mathbf{X}}(\mathbf{z}) d^m \mathbf{z},
 \end{aligned}$$

$U \in \mathcal{B}^n$ and $S \in \tilde{\mathcal{B}}^m$, where $\tilde{\mathcal{B}}^m$ is a restriction of \mathcal{B}^m to $V_{\mathbf{Z}}$ while

$$h(\mathbf{z}) \equiv \int_U \frac{f_{\mathbf{X}}(\mathbf{y}, \mathbf{z})}{f_{\mathbf{X}}(\mathbf{z})} d^n \mathbf{y}.$$

In (60), the first equality follows from the definition of $\tilde{\nu}_{\mathbf{1}_{\mathbf{Y}^{-1}(U)}}(S)$ (Definition 9), the third equality follows from the change of variables Theorem (Dudley (1989), § 4.1, p.92), while the last equality follows from Fubini's Theorem (Bartle (1966), Chapter 10, pp.119-120). Inserting (11) into the right-hand side of (10) yields, on the other hand,

$$\int_S \left[\int_U f_{\mathbf{X}}(\mathbf{y}|\mathbf{z}) d^n \mathbf{y} \right] f_{\mathbf{X}}(\mathbf{z}) d^m \mathbf{z} = \int_S k(\mathbf{z}) f_{\mathbf{X}}(\mathbf{z}) d^m \mathbf{z}.$$

Let $S_{1,2} \equiv \{\mathbf{z} : h(\mathbf{z}) \geq k(\mathbf{z})\}$. Then, the equality of $h(\mathbf{z})$ and $k(\mathbf{z})$ $Pr_{\mathbf{X}}^{\mathbb{Z}}$ -almost everywhere on $V_{\mathbf{Z}}$ follows immediately from Fatou's Lemma (see, for example, Bartle (1966), Chapter 4, Corollary 4.10 of Fatou's Lemma, pp.34-35), while the equality of $f_{\mathbf{X}}(\mathbf{y}, \mathbf{z})/f_{\mathbf{X}}(\mathbf{z})$ and $f_{\mathbf{X}}(\mathbf{y}|\mathbf{z})$ ν_L -almost everywhere on \mathbb{R}^n is obtained in an analogous way.

A.2 Proof of Lemma 1

Let $r : \mathbb{R} \times G \rightarrow \mathbb{R}$, $r(x, a) \equiv l(a^{-1}, x)$, be the *right action* of G on \mathbb{R} . Then, $r(x, a \circ b) = r[r(a, x), b]$ holds true for all $a, b \in G$ and for all $x \in \mathbb{R}$. A differentiation of $r(x, a \circ b)$ with respect to a thus yields

$$\partial_{a \circ b} r(x, a \circ b) \partial_a(a \circ b) = \partial_{r(x, a)} r[r(x, a), b] \partial_a r(x, a),$$

which for $b = a^{-1}$ reduces to

$$\begin{aligned}
 \partial_c r(x, c)|_{c=e} \partial_a(a \circ b)|_{b=a^{-1}} &= \\
 \partial_{r(x, a)} r[r(x, a), b]|_{b=a^{-1}} \partial_a r(x, a), &
 \end{aligned}$$

$c \equiv a \circ b$. The left-hand side of the above equation is zero due to the premise of the Lemma,

$$\partial_c r(x, c)|_{c=e} \equiv \partial_c l(c^{-1}, x)|_{c=e} = 0.$$

On the right-hand side, however, the first term,

$$\partial_{r(x, a)} r[r(x, a), b]|_{b=a^{-1}} = \partial_y l(a^{-1}, y) \equiv \partial_y g_{a^{-1}}(y)$$

is non-vanishing for all admissible values of the index a and for all real $y \equiv g_a(x)$ since differentiability of $l(a, x)$ with respect to x for every a is assumed. Then, $\partial_a r(x, a) \equiv \partial_a l(a^{-1}, x) = 0$ is implied for all permissible a , i.e., $g_a^{-1}(x)$ is permitted to depend on x only, say $g_a(x) \equiv h(x)$. When $g_e(x) = x$ is invoked, this further means $h(x) = x$ and the Lemma is proved.

A.3 Proof of Lemma 2

Suppose there exists a realization λ_0 of Λ for which the partial derivative (28) vanishes. Since the family I of direct distributions is invariant under \mathcal{G} , equation (29) applies which, when differentiated with respect to a and set afterwards $a = e$, yields

$$\begin{aligned}
 \partial_x F_I(x|\lambda) \partial_a l(a^{-1}, x)|_{a=e} \\
 = -\partial_\lambda F_I(x|\lambda) \partial_a \bar{l}(a^{-1}, \lambda)|_{a=e}.
 \end{aligned}$$

The second term on right-hand side of the above equation vanishes for $\lambda = \lambda_0$, which implies

$$\partial_a l(a^{-1}, x)|_{a=e} = 0 \quad ; \quad \forall x \in V_X.$$

This means, according to Lemma 1, that all transformations $g_a \in \mathcal{G}$ are trivial for all $x \in V_X(\lambda)$, which is in direct contradiction with the initial premises, so that the proof is completed.

A.4 Proof of Lemma 3

It is easily shown that every cdf $F_I(x|\lambda)$ of the form (34) solves (30). In order to demonstrate that the cdf's of the form (34) are also the only solutions of (30), suppose for a moment that $F_I(x|\lambda)$ can be written in terms of two independent variables, $H(x, \lambda)$ (31) and $K(x, \lambda) \equiv s(x) + \bar{s}(\lambda)$,

$$(61) \quad F_I(x|\lambda) = \Phi[H(x, \lambda), K(x, \lambda)] ,$$

where the functions $s(x)$ and $\bar{s}(\lambda)$ are defined *via* (32) and (33). Inserting (61) into (30) yields

$$\begin{aligned} \partial_K \Phi(H, K) [\partial_x H \partial_\lambda K - \partial_\lambda H \partial_x K] = \\ 2 s'(x) \bar{s}'(\lambda) \partial_K \Phi(H, K) = \\ 0 . \end{aligned}$$

Therefore, for $s'(x), \bar{s}'(\lambda) \neq 0$, $\partial_K \Phi(H, K)$ must vanish identically, such that the form (34) of $F_I(x|\lambda)$ is implied. If, on the other hand, any of $s'(x)$ and $\bar{s}'(\lambda)$ vanishes, $H(x, \lambda)$ and $K(x, \lambda)$ cease to be independent, i.e., $K(x, \lambda) = K[H(x, \lambda)]$, such that (34) again holds true, but since in this case $F_I(x|\lambda)$ is either a function of x alone, a function of λ alone, or a constant, such a solution is inadmissible for a cdf from a parametric family.

A.5 Proof of Proposition 3

According to the premises of the Proposition, a positive $f_I(\theta_1|\theta_2, \mathbf{x}_1, \mathbf{x}_2)$ exists and can be decomposed according to (41). Let $\theta_1' \in \tilde{V}_{\Theta_1}(\mathbf{x}_{1,2}, \theta_2)$ be another realization of Θ_1 fulfilling the conditions of the Proposition, such that

$$\begin{aligned} f_I(\theta_1'|\theta_2, \mathbf{x}_1, \mathbf{x}_2) &= \frac{f_I(\theta_1'|\theta_2, \mathbf{x}_2) f_I(\mathbf{x}_1|\theta_1', \theta_2)}{f_I(\mathbf{x}_1|\theta_2, \mathbf{x}_2)} \\ &= \frac{f_I(\theta_1'|\theta_2, \mathbf{x}_1) f_I(\mathbf{x}_2|\theta_1', \theta_2)}{f_I(\mathbf{x}_2|\theta_2, \mathbf{x}_1)} \end{aligned}$$

is also positive. Dividing the above equation with (41) yields

$$(62) \quad \frac{\kappa(\mathbf{x}_1, \theta_1', \theta_2)}{\kappa(\mathbf{x}_1, \theta_1, \theta_2)} = \frac{\kappa(\mathbf{x}_2, \theta_1', \theta_2)}{\kappa(\mathbf{x}_2, \theta_1, \theta_2)} ,$$

$\kappa(\mathbf{x}_{1,2}, \theta_1^{(')}, \theta_2) \equiv f_I(\theta_1^{(')}|\theta_2, \mathbf{x}_{1,2})/f_I(\mathbf{x}_{1,2}|\theta_1^{(')}, \theta_2)$. Clearly, in order to ensure equality in (62) for all \mathbf{x}_1 and \mathbf{x}_2 for which $f_I(\mathbf{x}_{1,2}|\theta_1^{(')}, \theta_2) > 0$, the left-hand and the right-hand side of the equation must

be independent of \mathbf{x}_1 and \mathbf{x}_2 , but may depend on θ_1, θ_1' , and θ_2 :

$$q(\theta_1, \theta_1', \theta_2) \equiv \frac{\kappa(\mathbf{x}_{1,2}, \theta_1', \theta_2)}{\kappa(\mathbf{x}_{1,2}, \theta_1, \theta_2)} .$$

The function $q(\theta_1, \theta_1', \theta_2)$ is factorizable,

$$\frac{\zeta_{I, \Theta_1|\theta_2}(\theta_1', \theta_2)}{\zeta_{I, \Theta_1|\theta_2}(\theta_1, \theta_2)} \equiv q(\theta_1, \theta_1', \theta_2) ,$$

such that

$$\begin{aligned} \eta_{I, \Theta_1|\theta_2}(\mathbf{x}_{1,2}, \theta_2) &\equiv \frac{\zeta_{I, \Theta_1|\theta_2}(\theta_1', \theta_2)}{\kappa(\mathbf{x}_{1,2}, \theta_1', \theta_2)} \\ &= \frac{\zeta_{I, \Theta_1|\theta_2}(\theta_1, \theta_2)}{\kappa(\mathbf{x}_{1,2}, \theta_1, \theta_2)} , \end{aligned}$$

which proves equation (43), while equation (44) is proved in a similar way by invoking (42) instead of (41).

A.6 Proof of Proposition 4

Suppose for a moment that a pdf for θ , $f_I(\theta|x)$, can be assigned to $\theta \in V_\Theta$ based on $x \in V_X - \tilde{V}_X$ for which partial derivative (27) vanishes. Since the family I of direct probability distributions is \mathcal{G} -invariant, the distributions assigned to Θ are invariant under the induced group $\tilde{\mathcal{G}}$ such that equation (51) applies. When differentiated with respect to a and set afterwards $a = e$, (51) further implies

$$\partial_x F_I(\theta|x) \partial_a l(a^{-1}, x)|_{a=e} = -\partial_\theta F_I(\theta|x) \partial_a \bar{l}(a^{-1}, \theta)|_{a=e}$$

for all $\theta \in V_\Theta$. The left-hand side of the above equation vanishes due to the premises, adopted at the beginning of the proof. Since, by Lemma 2, the second term on the right-hand side does not vanish anywhere on V_Θ , $\partial_\theta F_I(\theta|x) = f_I(\theta|x)$ must vanish for all $\theta \in V_\Theta$, which is incompatible with the normalization requirement (5). Therefore, the assumed existence of $f_I(\theta|x)$, based on x with vanishing derivative (27), inevitably leads to inconsistencies and is thus ruled out.

A.7 Proof of Lemma 4

Equation (52) holds true for all $\mu, a_1 \in \mathbb{R}$ and for all $\sigma, a_2 \in \mathbb{R}^+$. For $a_1 = \mu$ and $a_2 = \sigma$ we obtain $h(\mu, \sigma) = \zeta_{I, \Theta_1|\sigma}(\mu, \sigma)/\zeta_{I, \Theta_1|\sigma}(0, 1)$, while setting $a_1 = \mu$ and $a_2 = 1$ reveals factorizability of

$\zeta_{I,\Theta_1|\sigma}(\mu, \sigma)$:

$$(63) \quad \zeta_{I,\Theta_1|\sigma}(\mu, \sigma) = \frac{\zeta_{I,\Theta_1|\sigma}(\mu, 1) \zeta_{I,\Theta_1|\sigma}(0, \sigma)}{\zeta_{I,\Theta_1|\sigma}(0, 1)}.$$

By taking these findings into account, equation (52) reduces to

$$\begin{aligned} & \zeta_{I,\Theta_1|\sigma}(\mu, 1) \zeta_{I,\Theta_1|\sigma}(0, \sigma) [\zeta_{I,\Theta_1|\sigma}(0, 1)]^2 = \\ & \zeta_{I,\Theta_1|\sigma}(a_1, 1) \zeta_{I,\Theta_1|\sigma}(0, a_2) \times \\ & \zeta_{I,\Theta_1|\sigma}[(\mu - a_1)/a_2, 1] \zeta_{I,\Theta_1|\sigma}(0, \sigma/a_2), \end{aligned}$$

which for $a_1 = 0$ and $a_2 = \sigma$ yields $\zeta_{I,\Theta_1|\sigma}(\mu, 1) = \zeta_{I,\Theta_1|\sigma}(\mu/\sigma, 1)$. Hence, $\zeta_{I,\Theta_1|\sigma}(\mu, 1)$ must be a constant, such that, according to (63), $\zeta_{I,\Theta_1|\sigma}(\mu, \sigma)$ is a function of σ alone.

REFERENCES

- Bartle, R. G. (1966). *The Elements of Integration*. John Wiley & Sons.
- Bayes, R. T. (1763). An Essay towards solving a Problem in the Doctrine of Chances. *Philos. Trans. R. Soc. London*, **53**:370–418.
- Berger, J. O. (1980). *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag.
- Bernardo, J. M. (1979). Reference Posterior Distributions for Bayesian Inference. *J. R. Statist. Soc.*, **B 41**:113–147.
- Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Duxbury Press, 2nd edition.
- Chang, T. and Villegas, C. (1986). On a Theorem of Stein Relating Bayesian and Classical Inferences in Group Models. *Canad. J. Statist.*, **14**:289–296.
- Cox, R. T. (1946). Probability, Frequency and Reasonable Expectation. *Amer. J. Phys.*, **14**:1–13.
- Dawid, A. P., Stone, M., and Zidek, J. V. (1973). Marginalization Paradoxes in Bayesian and Structural Inference. *J. R. Statist. Soc.*, **B 35**:189–233.
- de Finetti, B. (1931). Sul significato soggettivo della probabilità. *Fund. Math.*, **17**:298–329.
- de Finetti, B. (1937). La prévision: ses lois logiques, ses sources objectives. *Ann. Inst. H. Poincaré*, **7**:1–68.
- Dudley, R. M. (1989). *Real Analysis and Probability*. Belmont, California: Wadsworth, Inc.
- Eaton, M. L. (1989). *Group Invariance Applications in Statistics*. Regional Conference Series in Probability and Statistics, Vol. 1. Published by the Institute of Mathematical Statistics and the American Statistical Association.
- Eaton, M. L. and Sudderth, W. D. (1993). Prediction in a multivariate normal setting: Coherence and incoherence. *Sankhyā*, **55**:481–493.
- Eaton, M. L. and Sudderth, W. D. (1999). Consistency and strong inconsistency of group-invariant predictive inferences. *Bernoulli*, **5**:833–854.
- Eaton, M. L. and Sudderth, W. D. (2002). Group invariant inference and right Haar measure. *J. Statist. Plann. Infer.*, **103**:87–99.
- Ferguson, T. S. (1967). *Mathematical Statistics – A Decision Theoretical Approach*. Academic Press.
- Fisher, R. A. (1935). The fiducial argument in statistical inference. *Ann. Eugen.*, **6**:391–398.
- Fisher, R. A. (1956). *Statistical Methods and Scientific Inference*. Edinburgh: Oliver & Boyd.
- Harney, H. L. (2003). *Bayesian Inference – Parameter Estimation and Decisions*. Springer.
- Hartigan, J. A. (1964). Invariant Prior Distributions. *Ann. Math. Statist.*, **35**:836–845.
- Hogg, R. V., McKean, J. W., and Craig, A. T. (2005). *Introduction to Mathematical Statistics*. Upper Saddle River, New Jersey: Pearson Prentice Hall, 6th edition.
- Jaynes, E. T. (1968). Prior Probabilities. *IEEE Trans. Syst. Sci. Cyber.*, **4**:227–241.
- Jaynes, E. T. (2003). *Probability Theory – The Logic of Science*. Cambridge University Press.
- Jeffreys, H. (1946). An invariant form of the prior probability in estimation problem. *Proc. R. Soc. London*, **A 186**:453–461.
- Jeffreys, H. (1957). *Scientific Inference*. Cambridge University Press, 2nd edition.
- Jeffreys, H. (1961). *Theory of Probability*. Oxford: Clarendon Press, 3rd edition.
- Kass, R. E. and Wasserman, L. (1996). The Selection of Prior Distributions by Formal Rules. *J. Amer. Statist. Assoc.*, **91**:1343–1370.
- Kemeny, J. G. (1955). Fair Bets and Inductive Probabilities. *J. Symb. Log.*, **20**:263–273.
- Kendall, M. G. (1949). On the reconciliation of theories of probability. *Biometrika*, **36**:101–116.
- Kolmogorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 2, Nr. 3. Berlin: Springer.
- Lad, F. (1996). *Operational Subjective Statistical Methods*. John Wiley & Sons, Inc.
- Laplace, P. S. (1774). Mémoire sur la probabilité des causes par les évènements. *Mem. Acad. R. Sci. Paris*, **6**:621–656.
- Laplace, P. S. (1886). *Œuvres Complètes - Tome Septième: Théorie Analytique des Probabilités*. Paris: Gauthier-Villars.
- Lehmann, E. L. (1986). *Testing Statistical Hypotheses*. John Wiley & Sons, Inc., 2nd edition.
- Lindley, D. V. (1958). Fiducial Distributions and Bayes' Theorem. *J. R. Statist. Soc.*, **B 20**:102–107.
- Nachbin, L. (1965). *The Haar Integral*. Princeton: Van Nostrand.
- Neyman, J. (1937). Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability. *Phil. Trans. R. Soc. London*, **A 236**:333–380.
- O'Hagan, A. (1994). *Bayesian Inference*. Kendall's Advanced Theory of Statistics, Vol. 2B. London: Arnold.
- Paris, J. B. (1994). *The Uncertain Reasoner's Companion – A Mathematical Perspective*. Cambridge University Press.
- Ramsey, F. P. (1931). *Foundations of Mathematics and other Essays*. London: Kegan, Paul, Trench, Trubner & Co.
- Rao, M. M. (1993). *Conditional Measures and Applications*. Marcel Dekker, Inc.
- Rényi, A. (1955). On a new axiomatic theory of probability.

- Acta Math. Acad. Sci. Hung.*, **6**:285–335.
- Robert, C. P. (2001). *The Bayesian Choice*. Springer, 2nd edition.
- Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell Sys. Tech. J.*, **27**:379–423 and 623–656.
- Shao, J. (1999). *Mathematical Statistics*. Springer-Verlag.
- Shimony, A. (1955). Coherence and the Axioms of Confirmation. *J. Symb. Log.*, **20**:1–28.
- Stein, C. (1965). Approximation of Improper Prior Measures by Prior Probability Measures. In LeCam, L. M. and Neyman, J., editors, *Bernoulli–Bayes–Laplace Anniversary Volume: Proceedings of an International Research Seminar. Statistical Laboratory, University of California, Berkeley, 1963*, pages 217–240. Springer-Verlag.
- Stein, E. M. and Shakarachi, R. (2005). *Princeton Lectures in Analysis, Vol. III – Real Analysis: Measure Theory, Integration, and Hilbert Spaces*. Princeton and Oxford: Princeton University Press.
- Stuart, A. and Ord, J. K. (2000). *Distribution Theory*. Kendall’s Advanced Theory of Statistics, Vol. 1. London: Arnold.
- Stuart, A., Ord, J. K., and Arnold, S. (1999). *Classical Inference and the Linear Model*. Kendall’s Advanced Theory of Statistics, Vol. 2A. London: Arnold.
- Villegas, C. (1977). Inner Statistical Inference. *J. Amer. Statist. Assoc.*, **72**:453–458.
- Villegas, C. (1981). Inner Statistical Inference II. *Ann. Stat.*, **9**:768–776.
- Wigner, E. P. (1959). *Group Theory and Its Applications to the Quantum Mechanics of Atomic Spectra*. Academic Press.